Nonlinear Dyn
https://doi.org/10.1007/s11071-025-11235-8

RESEARCH

Fixed-time concurrent learning-based robust approximate optimal control

Junkai Tan · Shuangsi Xue · Tiansen Niu · Kai Qu · Hui Cao · Badong Chen

Received: 24 February 2025 / Accepted: 9 April 2025

Abstract In this paper, we investigate a fixed-time concurrent learning-based actor-critic-identifier (FxT-CL-ACI) control scheme for approximating the optimal tracking controller and identifying uncertain system parameters online. The proposed FxT-CL-ACI control scheme is applied to solve the robust optimal tracking control problem for uncertain nonlinear systems with disturbances and actuator saturation. The interaction between the leader and follower in the Stackelberg game is modeled to achieve robust optimal tracking control with sequential optimization of H_2 and H_∞ performance indices. The effectiveness of the proposed FxT-CL-ACI control scheme is demonstrated by a numerical simulation and a hardware experiment on a UAV system. The results show that the FxT-CL-ACI control scheme can achieve robust optimal tracking control with fixed-time convergence and disturbance rejection, even in the presence of actuator saturation and uncertain system parameters.

J. Tan · S. Xue (✉) · T. Niu · K. Qu · H. Cao
Shaanxi Key Laboratory of Smart Grid, State Key Laboratory of Electrical Insulation and Power Equipment, and School of Electrical Engineering, Xi'an Jiaotong University, Xi'an 710049, China
e-mail: xssxjtu@xjtu.edu.cn

B. Chen
National Key Laboratory of Human–Machine Hybrid Augmented Intelligence, National Engineering Research Center for Visual Information and Applications, and the Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University, Xi'an 710049, China

Keywords Fixed-time concurent learning · Stackelberg game · Approximate optimal control · Neural network · Actor-critic-identifier

1 Introduction

Optimal control design is a fundamental problem in control theory with applications in various fields including robotics [1], aerospace [2], and autonomous systems [3]. In practical control systems, achieving optimal tracking performance while handling uncertainties [4], disturbances [5], and actuator constraints [6] remains a significant challenge. The presence of unknown dynamics and external disturbances can severely degrade control performance or even destabilize the system [7]. Although various robust and adaptive control methods have been developed, simultaneously optimizing tracking performance while ensuring predictable convergence time has not been fully addressed, particularly for nonlinear systems with both parametric uncertainties and input saturation. Stackelberg game theory provides a promising framework for solving such problems by modeling the interaction between the leader and follower in a hierarchical manner [8–10], in which the leader optimizes a performance index while anticipating the follower's response. Compared with other game theory-based control methods, such as non-zero sum cooperative game [11,12] and zero-sum game [13,14], Stackelberg games offer a more structured approach to solving optimal control problems with sequential optimization of performance indices, including non-zero sum games [15,16], and mixed H_2/H_∞ norms [17].

Traditional optimal control methods often rely on accurate system models and are sensitive to uncertainties [18,19]. While adaptive and robust control techniques can handle uncertainties and disturbances, they typically do not guarantee optimal performance [20,21]. Recent advances in reinforcement learning (RL), especially actor-critic frameworks, have enabled data-driven approaches for optimal control synthesis [22–24]. However, most existing actor-critic methods have two key limitations: the first is the lack of performance guarantees of convergence time [25,26], which may lead to unpredictable control performance in safety-critical applications, and the second is the inability to handle parameter uncertainties and external disturbances [27,28], which can significantly degrade control performance. Recent works have shown promise in handling input saturation and disturbances using RL-based control methods [29,30]. However, existing methods often struggle with the dual challenges of input saturation and parametric uncertainties, particularly when aiming for optimal performance [31,32]. Input saturation can severely degrade control performance and even destabilize the system if not properly addressed [33,34]. At the same time, unknown or uncertain system parameters make it difficult to synthesize optimal control policies that respect input constraints [35,36]. It is valuable to develop new control strategies that can jointly tackle these challenges and provide guaranteed performance and robustness.

Despite recent advances in optimal control and RL, while various robust and adaptive control methods have been proposed [37–39], ensuring fixed-time convergence has not been adequately addressed for the optimal control of nonlinear systems with parametric uncertainties and input saturation [40,41]. The lack of systematic frameworks that can jointly tackle these challenges motivates the development of new control strategies. Fixed-time stability theory has emerged as a promising solution by providing convergence guarantees within fixed time [42,43]. This property is particularly valuable for safety-critical applications requiring predictable performance. Papers [44,45] have shown that fixed-time learning methods can be applied to nonlinear systems with disturbances and uncertainties. The finite-time adaptive dynamic programming (ADP) method in [46,47] has demonstrated the effectiveness of finite-time learning for optimal control synthesis. While fixed-time control has been successfully applied to various control problems including stabilization and tracking, existing methods primarily focus on linear systems or systems with known dynamics [48,49], and its application to optimal tracking control for uncertain nonlinear systems remains largely unexplored. Also, complex interactions between learning-based optimal control and fixed-time stability have not been fully investigated.

Key challenges in optimal tracking control include achieving fixed-time convergence while ensuring robust performance against uncertainties and input constraints. Although finite-time control [46,49] has been studied extensively, guaranteeing fixed-time convergence independent of initial conditions remains challenging [47,48], especially for nonlinear systems with both parametric uncertainties and input saturation [28,40]. While our approach builds upon existing methods, the key innovation lies in the systematic integration of these techniques within a unified mathematical framework that provides fixed-time stability guarantees for Stackelberg game equilibria. This integration is non-trivial as it resolves fundamental theoretical conflicts between the asymptotic nature of traditional game-theoretic solutions and the fixed-time requirement of real-time control applications.

The main contributions are:

1. A fixed-time concurrent learning-based robust actor-critic-identifier (FxT-CL-ACI) control scheme is proposed to approximate the optimal tracking controller for uncertain nonlinear systems with disturbances and actuator saturation, where a FxT-CL mechanism with experience replay buffers is developed for training the ACI. Theoretical analysis proves that ACI weight errors converge to bounded regions in fixed time, an improvement over standard concurrent learning approaches such as [50,51] where asymptotic convergence is guaranteed, or recent works [4,23,47,52] that utilize experience replay but lack joint system identification with optimal control.
2. A Stackelberg game structure is developed to achieve robust optimal tracking control through sequential optimization with disturbance: The controller act as leader pursues H_2 performance by minimizing tracking error and control energy, while the disturbance act as follower optimizes H_∞ per-

Journal: 11071 MS: 11235 TYPESET DISK LE CP Disp.:2025/5/2 Pages: 21 Layout: Medium

formance to ensure disturbance attenuation. This structure balances trade-offs between performance and robustness compared with [8,9,11].

3. A hardware experiment using a UAV platform provides comprehensive validation of the proposed approach. Results validate both robust tracking performance and fixed-time convergence properties in a practical setting.

Paper organization: Sect. 2 covers nonlinear system tracking and fixed-time control. Section 3 describes robust optimal tracking control with Stackelberg game framework. Section 4 presents the FxT-CL-ACI control scheme. Sections 5–6 provide simulation and UAV experimental validation. Section 7 summarizes findings and future work.

Notations: $\mathbb{R}^n$ and $\mathbb{R}^{n\times m}$ denote the n-dimensional Euclidean space and space of $n \times m$ real matrices; $\|\cdot\|$ denotes Euclidean norm; $\mathrm{diag}([a_1, ..., a_n])$ denotes a diagonal matrix with elements a_i; $\mathrm{sat}(u)$ denotes vector saturation; $\mathrm{sat}_{\mu_i}(u_i)$ denotes component-wise saturation with bound μ_i; $\mathrm{sig}(\cdot)$, $\mathrm{sign}(\cdot)$ and $\tanh(\cdot)$ denote sign, signum and hyperbolic tangent functions, with $\mathrm{sig}^{\gamma}(\cdot) = |\cdot|^{\gamma}\,\mathrm{sign}(\cdot)$; $\bar{\lambda}(\cdot)$ and $\underline{\lambda}(\cdot)$ denote maximum and minimum eigenvalues of a matrix.

2 Preliminaries

2.1 Nonlinear system with disturbances and saturation

Consider the following continuous-time nonlinear system with disturbances and actuator saturation:

$$\begin{aligned}
&\dot{x}(t) = f(x(t)) + g(x(t))\mathrm{sat}(u(t)) + k(x(t))\omega(t)\\
&\mathrm{sat}(u(t)) = [\mathrm{sat}_{\mu_1}(u_1), ..., \mathrm{sat}_{\mu_m}(u_m)]^{\top}\\
&\mathrm{sat}_{\mu_i}(u_i) = \begin{cases} \mu_i, & u_i > \mu_i \\ u_i, & |u_i| \le \mu_i \\ -\mu_i, & u_i < -\mu_i \end{cases}, \quad i = 1, ..., m
\end{aligned} \tag{1}$$

where $x(t) \in \mathbb{R}^n$ is the system state vector, $f : \mathbb{R}^n \to \mathbb{R}^n$ represents the unknown drift dynamics, $g : \mathbb{R}^n \to \mathbb{R}^{n\times m}$ is the input matrix, $k : \mathbb{R}^n \to \mathbb{R}^{n\times m}$ is the disturbance matrix, $u(t) \in \mathbb{R}^m$ is the control input subject to saturation with bounds $\mu_i > 0$, and $\omega(t) \in \mathbb{R}^m$ denotes external disturbances. The functions $f(x)$, $g(x)$ and $k(x)$ are assumed to be locally Lipschitz continuous. Let $x_d(t) \in \mathbb{R}^n$ denote the desired trajectory, which may be time-varying (e.g., $x_d = \sin(wt)$) and is governed by:

$$\dot{x}_d(t) = f_d(x_d(t), t) \tag{2}$$

where $f_d : \mathbb{R}^n \times \mathbb{R}^+ \to \mathbb{R}^n$ defines the reference dynamics with explicit time-dependency to accommodate periodic or other time-varying trajectories. Let $X(t) = x(t) - x_d(t)$ denote the tracking error between the actual state and desired trajectory. The tracking error dynamics can be derived as:

$$\begin{aligned}
\dot{X}(t) &= \dot{x}(t) - \dot{x}_d(t)\\
&= [f(x) - f_d(x_d)] + g(x)\mathrm{sat}(u) + k(x)\omega\\
&= F(X) + G(X)\mathrm{sat}(U) + K(X)\omega
\end{aligned} \tag{3}$$

where $F(X) = f(X + x_d) - f_d(x_d)$ represents the transformed drift dynamics, $G(X) = g(X + x_d)$ is the transformed input matrix, $K(X) = k(X + x_d)$ is the transformed disturbance matrix, and $U(t) = u(t)$ is the control input vector. The system (3) captures both the tracking objective and the effects of disturbances and input saturation. To ensure prescribed performance tracking with bounded error trajectories, we introduce the following performance transformation in the next subsection. Let $\Phi(U)$ denote the input cost function inspired by the hyperbolic tangent function [31,32]:

$$\Phi(U) = \int_0^u 2\mu R \phi^{-1}(\gamma_u/\mu)\,\mathrm{d}\gamma_u \tag{4}$$

where $u = [u_1, ..., u_m]^{\top}$ represents the input vector, $\phi(\cdot) = \tanh(\cdot)$ is the hyperbolic tangent activation function for smooth approximation of saturation, $\phi^{-1}(\cdot)$ is the inverse hyperbolic tangent function, $\mu_i > 0$ are the component-wise saturation bounds defined in (1), $\mu = \mathrm{diag}([\mu_1, ..., \mu_m])$ denotes the diagonal matrix of saturation bounds, $R = \mathrm{diag}([r_1, ..., r_m])$ is the positive definite input weighting matrix, and γ_u is the dummy integration variable. The input cost $\Phi(U)$ provides a smooth penalty for control inputs approaching saturation limits. The following assumption characterizes the system dynamics and cost functions.

Assumption 2.1 (Boundedness and Continuity [1,5]) The following conditions hold for system (1):

1. *Bounded Disturbances:* The disturbance matrix $K(X)$ is continuous and bounded by: $\|K(X)\| \le K_H$ for all $X \in \chi$.

Journal: 11071 MS: 11235 TYPESET DISK LE CP Disp.:2025/5/2 Pages: 21 Layout: Medium

2. *Continuity and Differentiability:* On a compact set $X \in \chi \subset \mathbb{R}^n$, the drift dynamics $F(X)$ and input matrix $G(X)$ are continuously differentiable with $F(0) = 0$. Furthermore, there exist positive constants L_F, L_G, and G_H such that $\|F(X_1) - F(X_2)\| \leq L_F\|X_1 - X_2\|$, $\|G(X_1) - G(X_2)\| \leq L_G\|X_1 - X_2\|$, and $\|G(X)\| \leq G_H$ for all $X, X_1, X_2 \in \chi$.
3. *Bounded Input Cost:* The cost matrices Q and R are positive definite symmetric matrices satisfying $0 < \underline{\lambda}_Q \mathscr{I} \leq Q \leq \bar{\lambda}_Q \mathscr{I}$, and $0 < \underline{\lambda}_R \mathscr{I} \leq R \leq \bar{\lambda}_R \mathscr{I}$, where $\underline{\lambda}_Q, \bar{\lambda}_Q, \underline{\lambda}_R, \bar{\lambda}_R$ are positive constants.

2.2 Fixed-time stability

To achieve fixed-time stabilization of the system states, we introduce key definitions and lemmas from fixed-time stability theory that form the foundation of our approach.

Definition 1 (Fixed-time Stability [6,36]) For system (1), if there exists a settling time $T > 0$ independent of initial conditions, such that:

$$V(x(t)) \leq \begin{cases} \beta(V(x(0)), t), & \text{if } 0 \leq t < T \\ \epsilon, & \text{if } t \geq T \end{cases} \tag{5}$$

where $\epsilon > 0$ is a small positive constant representing the terminal bound, such that $\|x(T) - x^*\| \leq \delta$ for some small $\delta > 0$. The system is then called fixed-time stable, with the equilibrium point x^* being reached within a fixed time T up to a small bounded error δ.

Note that in practical implementations, exact convergence to x^* at exactly time T may not be achievable due to numerical limitations and approximation errors. The above definition acknowledges that the system state converges to a small neighborhood of the equilibrium point rather than exactly to x^*. To achieve fixed-time convergence, we propose the following fractional power transformation:

$$\varXi(x, x^*) = \frac{V(x, x^*)^{\gamma_1}}{1 - \gamma_1} + \frac{V(x, x^*)^{\gamma_2}}{1 - \gamma_2} \tag{6}$$

where $V(x, x^*)$ is the original function, $\gamma_1 \in (0, 1)$ and $\gamma_2 > 1$ are fractional exponents selected to ensure fixed-time stability.

Lemma 2.2 *(Fixed-time Convergence [37,38]) Consider system (1) with the transformed function (6). If the time derivative satisfies:*

$$\dot{\varXi} \leq -k_1 \varXi^{\gamma_1} - k_2 \varXi^{\gamma_2} \tag{7}$$

where $k_1, k_2 > 0$, then the system converges to equilibrium in fixed time bounded by:

$$T \leq \frac{1}{k_1(1 - \gamma_1)} + \frac{1}{k_2(\gamma_2 - 1)} \tag{8}$$

The transformed value function (6) with fractional powers enables fixed-time convergence independent of initial conditions. This transformation will be utilized in developing fixed-time concurrent learning algorithm in Sect. 4.

3 Problem formulation: robust optimal tracking control with input saturation

Considering the nonlinear system (1) with disturbances and actuator saturation, the objective is to design a robust optimal tracking controller that achieves fixed-time convergence and disturbance rejection. The following problem formulation establishes the Stackelberg game framework for solving the robust optimal tracking control problem. First, we define finite L_2 gain stability required for robust control design.

Definition 2 (Finite L_2-gain stable [22,53]) For the nonlinear system (1), if there exists a positive constant γ such that for any bounded disturbance input $\omega(t)$, the output $y(t)$ is bounded and satisfies:

$$\int_0^\infty \|y(t)\|^2 dt \leq \gamma^2 \int_0^\infty \|\omega\|^2 dt \tag{9}$$

where $y(t) = [\sqrt{Q}X(t), \sqrt{R}U(t)]^\top$ is the output vector, then the system is finite L_2-gain stable with disturbance attenuation level γ. This stability criterion corresponds to the H_∞ norm of the closed-loop system, measuring the maximum energy gain from disturbances to regulated outputs. The parameter $\gamma > 0$ is the prescribed disturbance attenuation level. If the closed-loop dynamics is stable with a minimum gain $\gamma^* > 0$, it remains stable with any $\gamma > \gamma^*$ [22,53].

For optimal control design, we define the following H_2 and H_∞ performance indices with the control input

Journal: 11071 MS: 11235 TYPESET DISK LE CP Disp.:2025/5/2 Pages: 21 Layout: Medium

Table 1 Stackelberg game framework for robust optimal control

Game Level	Stackelberg Game Description
Level 1: *Leader Optimization*	The leader pursues H_2 optimal performance by minimizing: $J_1^*(X_0) = \min_{U\in\Omega_U} J_1(X_0, U, \omega^*)$
Level 2: *Follower Response*	Given leader's strategy U^*, follower optimizes H_∞ performance: $J_2^*(X_0) = \min_{\omega\in\Omega_W} J_2(X_0, U^*, \omega)$
Level 3: *Stackelberg Equilibrium*	The game reaches equilibrium when: $\begin{cases} U^* = \arg\min\limits_{U\in\Omega_U} J_1(X_0, U, \omega^*) \\ \omega^* = \arg\min\limits_{\omega\in\Omega_W} J_2(X_0, U^*, \omega) \end{cases}$

cost function $\Phi(U)$ in (4):

$$J_1(X_0, U, \omega) = \int_t^\infty \Big(X^\top QX + \Phi(U)\Big)\, d\tau \tag{9}$$

$$J_2(X_0, U, \omega) = \int_t^\infty \Big(\gamma^2\|\omega\|^2 - X^\top QX - \Phi(U)\Big)\, d\tau \tag{10}$$

where $Q = \mathrm{diag}([q_1, ..., q_n])$ is the positive definite state weighting matrix, $\gamma > 0$ is the disturbance attenuation level, J_1 measures H_2 performance and J_2 measures H_∞ performance. The robust optimal tracking control problem is formulated as follows:

Problem 1 Design a Stackelberg game-based controller for system (1) that:

1. Achieves optimal control and worst-case disturbance rejection with fixed-time convergence via:
$$\begin{cases} U^*(t) = \arg\min\limits_{U\in\Omega_U} J_1(X_0, U, \omega^*) \\ \omega^*(t) = \arg\min\limits_{\omega\in\Omega_W} J_2(X_0, U^*, \omega) \end{cases} \tag{11}$$
where J_1 and J_2 are defined in (9)-(10)
2. Ensures closed-loop finite L_2-gain stability per (9) with fixed-time convergence.

The Stackelberg game framework for solving this problem is shown in Table 1, which establishes a three-level hierarchical structure between the leader (controller) and follower (disturbance).

To solve the robust optimal tracking control problem, a Stackelberg game framework is established as shown in Table 1. First, we define the Stackelberg game formally:

Definition 3 (Stackelberg Game [8]) Consider a two-player game with:

– A leader $\mathbb{L}$ pursuing H_2 performance index J_1 in (9)
– A follower $\mathbb{F}$ pursuing H_∞ performance index J_2 in (10)

The game proceeds as follows:

1. The leader commits to a control strategy $U \in \Omega_U$ without knowing follower's choice
2. The follower observes leader's strategy U and responds with disturbance $\omega^*(U)$ that solves:
$$J_2(X_0, U, \omega^*(U)) = \min_{\omega\in\Omega_W} J_2(X_0, U, \omega) \tag{12}$$
3. Anticipating follower's response $\omega^*(U)$, the leader chooses optimal U^* that solves:
$$J_1(X_0, U^*, \omega^*(U^*)) = \min_{U\in\Omega_U} J_1(X_0, U, \omega^*(U)) \tag{13}$$

The resulting pair $(U^*, \omega^*(U^*))$ forms the Stackelberg equilibrium.

Remark 1 (Stackelberg vs. Nash Equilibrium) Unlike Nash equilibrium where players decide simultaneously, our approach uses Stackelberg equilibrium (Table 1) with sequential decisions. The controller (leader) acts first, anticipating the disturbance (follower) response. This hierarchical structure provides stronger performance guarantees than Nash solutions [41], creating an effective framework for balancing nominal performance and disturbance rejection [17]. Our fixed-time learning method ensures convergence to Stackelberg equilibrium within bounded time.

Based on the Stackelberg game definition, the follower pursues H_∞ performance by optimizing the value function J_2^* defined by the following minimization problem:

$$\begin{aligned} J_2^* &= \min_\omega J_2(X_0, U, \omega) \\ &= \min_\omega \int_t^\infty \Big(\gamma^2\|\omega\|^2 - X^\top QX - \Phi(U)\Big)\, d\tau \end{aligned} \tag{14}$$

Journal: 11071 MS: 11235 TYPESET DISK LE CP Disp.:2025/5/2 Pages: 21 Layout: Medium

The corresponded follower's Hamiltonian could be defined as:

$$H_{\mathbb{F}} =\nabla J_2^{*\top}(F+GU+K\omega)+\gamma^2\|\omega\|^2 - X^\top QX-\Phi(U) \tag{15}$$

By taking derivative of $H_{\mathbb{F}}$ (15) with respect to ω, the follower's optimal disturbance can be obtained as:

$$\omega^*(U) = -\frac{1}{2\gamma^2}K^\top\nabla J_2^* \tag{16}$$

To solve the Stackelberg game, a costate λ_2 is introduced to capture the follower's response to the leader's control strategy. Inspires by literature [9,17], the follower's costate λ_2 is defined as $\dot{\lambda}_2 = -\nabla H_{\mathbb{F}}$, where $\nabla H_{\mathbb{F}}$ is the gradient of the follower's Hamiltonian $H_{\mathbb{F}}$ with respect to X. For the leader pursuing H_2 performance, derived from original optimal value function (9), the revised optimal value function J_1^* incorporating the follower's costate λ_2 is defined as:

$$\begin{aligned} J_1^* &= \min_U J_1(X_0,U,\omega^*) \\ &= \min_U \int_t^\infty \left(X^\top QX+\Phi(U)+\eta^\top\dot{\lambda}_2\right)d\tau \end{aligned} \tag{17}$$

Then the corresponding Hamiltonian of (17) for the leader is derived as:

$$H_{\mathbb{L}} = \nabla J_1^{*\top}(F+GU+K\omega) + \eta^\top\dot{\lambda}_2+X^\top QX+\Phi(U) \tag{18}$$

where η is the Lagrange multiplier associated with the follower's costate λ_2, the Lagrange multiplier dynamics is given by $\dot{\eta}(t) = -\nabla_{\nabla J_2^*}H_{\mathbb{L}}$, $\nabla_{\nabla J_2^*}$ denotes the gradient with respect to ∇J_2^*. Minimizing $H_{\mathbb{L}}$ with respect to U yields the leader's optimal control:

$$U^*(\omega^*) = -\mu\phi\left(\frac{R^{-1}}{2\mu}\left(G^\top\nabla J_1^* - \nabla J_2^{*\top}\nabla G\eta\right)\right) \tag{19}$$

where ∇G is the gradient of the input matrix G with respect to X. In summary, the optimal control policies for the Stackelberg game are:

$$\begin{cases} U^* = -\mu\phi\left(\frac{R^{-1}}{2\mu}\left(G^\top\nabla J_1^* - \nabla J_2^{*\top}\nabla G\eta\right)\right) \\ \omega^* = -\frac{1}{2\gamma^2}K^\top\nabla J_2^* \end{cases} \tag{20}$$

Unlike approaches that handle unknown system internal dynamics without explicit identification [54], our framework includes system identification to achieve enhanced control performance and fixed-time guarantees. This design choice enables precise coordination in UAV tracking scenarios and provides mathematical tractability for establishing comprehensive fixed-time stability proofs.

Remark 2 (Symmetrical Saturation Constraints) This paper employs symmetrical constraints that align with our UAV platform's actuator characteristics ($\|V_{max}\| = 2m/s$). While recent research [55] has explored asymmetrical saturation models, symmetric constraints enable more elegant stability proofs within our fixed-time framework while still capturing essential constraint dynamics. Our continuous control approach provides smoother trajectory tracking with reduced mechanical jerk-a critical advantage for precision UAV control. Future work will extend our framework to asymmetrical constraints and potentially incorporate event-triggered mechanisms to balance computational efficiency with fixed-time guarantees.

Remark 3 (Stackelberg Game Structure) A Stackelberg game features sequential decision-making where a leader moves first, followed by responders who maximize their own benefits [17,28]. In our approach, we employ this mathematical structure as an optimization paradigm rather than describing physical UAV interactions. The controller (leader) and disturbance (follower) function as mathematical entities in a sequential optimization framework, with the controller anticipating the disturbance response. This formulation balances H_2 optimal performance (minimizing tracking error and control energy) with H_∞ robustness (disturbance attenuation) without requiring an actual leader-follower hierarchy between physical agents.

Due to the complexity of nonlinear dynamics and robust performance indices, obtaining explicit solutions for the optimal control inputs is challenging. Therefore, in the next section, we develop a RL-based approximation method using an actor-critic-identifier

Journal: 11071 MS: 11235 TYPESET DISK LE CP Disp.:2025/5/2 Pages: 21 Layout: Medium

structure to approximate the optimal value functions and control policies while identifying uncertain system parameters online.

4 Main results: fixed-time concurrent learning-based actor-critic-identifier

This section presents an actor-critic-identifier architecture to approximate the robust optimal tracking control solution. First, the optimal value functions and control inputs are reconstructed using actor-critic neural networks (NNs). Then, uncertain system parameters are identified online via an identifier. Finally, with the identified parameters and reconstructed optimal solutions, Bellman errors are established and minimized to train the actor-critic NNs.

4.1 Value function approximation via actor-critic

The optimal value functions for both leader and follower agents are approximated using critic neural networks:

$$J_i^*(X) = W_{ci}^\top \varphi_{ci}(X) + \delta_{ci}(X),\ i = 1, 2 \tag{21}$$

$$\nabla J_i^*(X) = \nabla\varphi_{ci}^\top(X) W_{ci} + \nabla\delta_{ci}^\top(X),\ i = 1, 2 \tag{22}$$

where $W_{ci} \in \mathbb{R}^{n_{\varphi_{ci}} \times 1}$ denotes the ideal critic NN weights, $\varphi_{ci}(X)$ represents the activation functions, and $\delta_{ci}(X)$ captures the reconstruction errors. For control policy approximation, actor neural networks are employed:

$$\begin{aligned} U^*(X) = &- \mu\phi\left(\frac{1}{2\mu}\left(R^{-1}G^\top\left(\nabla\varphi_{a1}^\top W_{a1} + \nabla\delta_{a1}^\top\right)\right.\right. \\ &\left.\left. - \left(W_{a2}^\top\nabla\varphi_{a2} + \nabla\delta_{a2}\right)\nabla G\eta\right)\right) \end{aligned} \tag{23}$$

$$\omega^*(X) = -\frac{K^\top}{2\gamma^2}\left(\nabla\varphi_{a2}^\top W_{a2} + \nabla\delta_{a2}^\top\right) \tag{24}$$

where $W_{ai} \in \mathbb{R}^{n_{\varphi_{ai}} \times 1}$ represents the ideal actor NN weights, and $\delta_{ai}(X)$ denotes the reconstruction errors. Since the ideal weights are unknown in practice, estimated weights are utilized:

$$\hat{J}_i(X) = \hat{W}_{ci}^\top \varphi_{ci}(X),\ i = 1, 2 \tag{25}$$

$$\begin{aligned} \hat{U}(X) = -\mu\phi\left(\frac{1}{2\mu}\left(R^{-1}G^\top\nabla\varphi_{a1}^\top\hat{W}_{a1}\right.\right. \\ \left.\left. -\hat{W}_{a2}^\top\nabla\varphi_{a2}\nabla G\eta\right)\right) \end{aligned} \tag{26}$$

$$\hat{\omega}(X) = -\frac{K^\top}{2\gamma^2}\nabla\varphi_{a2}^\top\hat{W}_{a2} \tag{27}$$

where $\hat{W}_{ci}$ and $\hat{W}_{ai}$ denote the estimated NN weights.

Remark 4 (Structure of Hamiltonian Functions) Regarding the Hamiltonian function $H_{\mathbb{L}}$ in (18), it's worth clarifying the representation of the optimal control $U^*(x)$ and its gradient. In (20), the optimal control depends on the value function gradients ∇J_1^* and ∇J_2^*, which are approximated by neural networks as $\nabla J_1^* \approx \nabla\varphi_{a1}^\top\hat{W}_{a1}$ and $\nabla J_2^* \approx \nabla\varphi_{a2}^\top\hat{W}_{a2}$. While the gradient of U^* ($\hat{U}$) with respect to state X is naturally captured in the actor-critic architecture through activation function gradients $\nabla\varphi_{ai}(X)$. This is reflected in the Bellman errors (32)-(33), where state derivatives are handled by the neural network structure. The concurrent learning approach ensures accurate approximation of both U^* and its gradient through experience replay buffers, which enhance learning by storing historical data samples.

Remark 5 (Selection of Activation Functions) The selection of activation functions in equations (26)-(27) is a critical design choice affecting both approximation accuracy and computational efficiency. For general nonlinear systems, activation functions should not only match the complexity of the underlying optimal control solution, but also maintain differentiability for stable gradient-based learning and rain computational efficiency. In this work, we selected fractional power activation functions with $[X_1^{\alpha+1}, \ldots, (X_1X_2)^{\alpha+1}]^\top$ because they satisfy these requirements while enhancing approximation capability for nonlinear optimal control problems.

4.2 System identification via identifier

For systems with parametric uncertainties, the drift dynamics are parameterized as:

$$F(X) = W_\theta^\top\varphi_\theta(X) + \delta_\theta(X) \tag{28}$$

where $\varphi_\theta \in \mathbb{R}^p$ contains the basis functions, $W_\theta \in \mathbb{R}^{p\times n}$ represents the unknown parameters, and $\delta_\theta(X)$

Journal: 11071 MS: 11235 TYPESET DISK LE CP Disp.:2025/5/2 Pages: 21 Layout: Medium

denotes the approximation error. The estimated drift dynamics are given by:

$$\hat{F}(X) = \hat{W}_\theta^\top \varphi_\theta(X) \tag{29}$$

where $\hat{W}_\theta \in \mathbb{R}^p$ represents the estimated parameters. The identification error ε_θ is defined as:

$$\varepsilon_\theta = \mathscr{F}(X) - \hat{W}_\theta^\top \varphi_\theta(X) \tag{30}$$

where $\mathscr{F}(X)$ represents the measured true drift dynamics. Utilizing fixed-time concurrent learning, parameters are estimated online via:

$$\dot{\hat{W}}_\theta = \frac{\Gamma_\theta k_\theta}{N} \sum_{j=1}^{N} \varphi_\theta(X^j)[\mathrm{sig}^{\gamma_1}(\varepsilon_\theta^j) + \mathrm{sig}^{\gamma_2}(\varepsilon_\theta^j)] \tag{31}$$

where $\hat{W}_\theta \in \mathbb{R}^p$ represents estimated parameters, $\gamma_1 \in (0, 1)$ and $\gamma_2 > 1$ are fractional exponents, $\Gamma_\theta \in \mathbb{R}^{p\times p}$ is positive definite, k_θ denotes the learning rate, N indicates the experience replay buffer size, and $\mathrm{sig}(x)$ is the sign function. Based on the identified dynamics, the Bellman errors are formulated as:

$$\begin{aligned}\hat{\varepsilon}_1 &= (\nabla J_1^*)^\top (\hat{W}_\theta^\top \varphi_\theta + G\hat{U} + K\hat{\omega}) \\ &\quad + X^\top QX + \Phi(\hat{U}) + \eta^\top \dot{\lambda}_2 \end{aligned}\tag{32}$$

$$\begin{aligned}\hat{\varepsilon}_2 &= (\nabla J_2^*)^\top (\hat{W}_\theta^\top \varphi_\theta + G\hat{U} + K\hat{\omega}) \\ &\quad + \gamma^2\|\hat{\omega}\|^2 - X^\top QX - \Phi(\hat{U}) \end{aligned}\tag{33}$$

where $\hat{\varepsilon}_1$ and $\hat{\varepsilon}_2$ represent the Bellman errors for leader and follower agents respectively.

Remark 6 (Fractional Power Signum Function) To address potential confusion, we clarify that $\mathrm{sig}^\gamma(x)$ represents the fractional power signum function as: $\mathrm{sig}^\gamma(x) = |x|^\gamma \,\mathrm{sign}(x)$, where $\mathrm{sign}(x)$ is the standard signum function. This notation follows established literature in fixed-time stability theory [44,56]. For $\gamma > 1$, this function is continuous and differentiable everywhere, resulting in standard ODEs. For $0 < \gamma < 1$, while not differentiable at the origin, the function remains continuous, and the resulting differential equations have been rigorously shown to possess well-defined solutions in the Filippov sense [57]. The combination of terms with $\gamma_1 \in (0, 1)$ and $\gamma_2 > 1$ in our update laws enables the fixed-time convergence properties proven in subsection 4.4.

4.3 Fixed-time concurrent learning

In this subsection, we present the online weight update mechanism for the actor-critic neural networks based on minimizing Bellman errors. The learning process utilizes experience replay buffers to enhance convergence and stability.

Both leader and follower agents maintain historical experience replay buffers:

$$\begin{cases}\mathscr{D}_1(t) = \{\hat{U}(t), \hat{\varepsilon}_1(t), \{\hat{U}^j(t), \hat{\varepsilon}_1^j(t)\}_{j=1}^N\} \\ \mathscr{D}_2(t) = \{\hat{\omega}(t), \hat{\varepsilon}_2(t), \{\hat{\omega}^j(t), \hat{\varepsilon}_2^j(t)\}_{j=1}^N\}\end{cases}$$

where $\{\hat{U}^j(t), \hat{\varepsilon}_1^j(t)\}$ and $\{\hat{\omega}^j(t), \hat{\varepsilon}_2^j(t)\}$ represent historical data samples for leader and follower agents respectively. Additionally, the identifier maintains its own experience replay buffer:

$$\mathscr{D}_\theta(t) = \{\hat{F}(t), \varepsilon_\theta(t), \{\hat{F}^j(t), \varepsilon_\theta^j(t)\}_{j=1}^N\}$$

where $\{\varphi_\theta^j(t), \varepsilon_\theta^j(t)\}$ represent historical data samples.

The actor-critic weights are updated by minimizing the following fractional fixed-time Bellman errors:

$$\begin{aligned}E_i &= \|\hat{\varepsilon}_i\|^{\gamma_1+1} + \|\hat{\varepsilon}_i\|^{\gamma_2+1} \\ &\quad + \sum_{k=1}^{N}\left(\|\hat{\varepsilon}_i^k\|^{\gamma_1+1} + \|\hat{\varepsilon}_i^k\|^{\gamma_2+1}\right), \\ &\quad i = 1, 2\end{aligned}\tag{34}$$

The critic NN weights are updated using concurrent learning-based gradient descent:

$$\begin{aligned}\dot{\hat{W}}_{ci} &= -\Gamma_{ci} k_{ci,1} \rho_i [\mathrm{sig}^{\gamma_1}(\hat{\varepsilon}_i) + \mathrm{sig}^{\gamma_2}(\hat{\varepsilon}_i)] \\ &\quad - \frac{\Gamma_{ci} k_{ci,2}}{N} \sum_{k=1}^{N} \rho_i^k \\ &\quad [\mathrm{sig}^{\gamma_1}(\hat{\varepsilon}_i^k) + \mathrm{sig}^{\gamma_2}(\hat{\varepsilon}_i^k)],\ i = 1, 2\end{aligned}\tag{35}$$

where $k_{ci,j} > 0$ $(i = 1, 2;\ j = 1, 2)$ are learning rates, $\rho_i = \sigma_i/\left(\sigma_i^\top \sigma_i + 1\right)^2$ is the normalized regression vector, $\rho_i^k = \sigma_i^k/\left(\sigma_i^{k\top}\sigma_i^k + 1\right)^2$ is the historical normalized regression vector, $\sigma_i = \nabla\varphi_{ci}^\top(X)(\hat{W}_\theta^\top \varphi_\theta + G\hat{U} + K\hat{\omega})$ is the current regression vector, and $\sigma_i^k = \nabla\varphi_{ci}^\top(X^k)(\hat{W}_\theta^\top \varphi_\theta + G\hat{U}^k + K\hat{\omega}^k)$ is the historical regression vector.

Define the actor-critic NNs error as $\varepsilon_{ai} = \hat{W}_{ai} - \hat{W}_{ci}$. The actor NN weights are updated using gradient descent:

$$\begin{aligned}\dot{\hat{W}}_{ai} &= -\Gamma_{ai}\left[k_{ai,1}\,\mathrm{sig}^{\gamma_1}(\varepsilon_{ai})\right. \\ &\quad \left.+ k_{ai,2}\,\mathrm{sig}^{\gamma_2}(\varepsilon_{ai})\right]\ i = 1, 2\end{aligned}\tag{36}$$

Journal: 11071 MS: 11235 TYPESET DISK LE CP Disp.:2025/5/2 Pages: 21 Layout: Medium

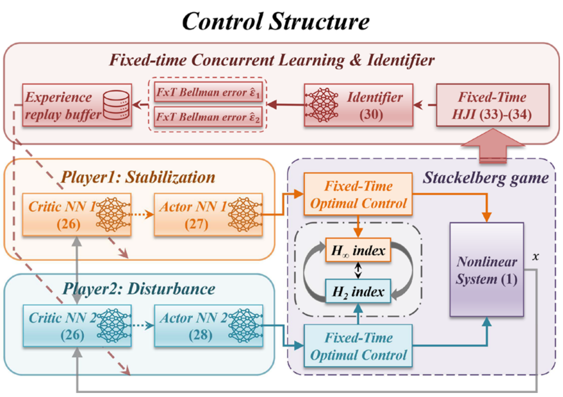


Fig. 1 Control structure of the FxT-CL-ACI control scheme

where $k_{ai,j} > 0$ $(i = 1, 2;\ j = 1, 2)$ are learning rates and $\Gamma_{ai} \in \mathbb{R}^{n_\varphi \times n_\varphi}$ are positive definite matrices.

The control structure of the proposed FxT-CL-ACI scheme is illustrated in Fig. 1. The scheme integrates: Actor-Critic-Identifier NNs approximating optimal value functions, control policies and uncertain system via equations (26), (27) and (29) with fixed-time guarantees; Experience Replay Buffers storing historical data samples to enhance learning stability; and FxT update mechanisms ensuring rapid convergence using fractional-power exponents $\gamma_1 \in (0, 1)$ and $\gamma_2 > 1$ as in (35) and (36). The architecture features bidirectional information flow where the leader optimizes H_2 performance while anticipating the follower's response optimizing H_∞ performance, creating a hierarchical optimization structure that balances tracking accuracy against disturbance rejection via Stackelberg equilibrium in (20). The detailed implementation is presented in Algorithm 1. To highlight these distinctions clearly, we compare our FxT-CL-ACI with existing methods in Table 2.

Remark 7 (Comparison with Other RL Methods) Unlike NN-based constrained RL in [27,34] that handles constraints but lacks fixed-time guarantees, our approach ensures predictable convergence times-critical for UAV applications. The integral-based actor-critic [33] addresses specific dynamics but is limited to single-agent optimization without our multi-objective framework. RISE-based RL [29] compensates for time-delays but requires symmetric uncertainty bounds and lacks game-theoretic optimization. Our FxT-CL-ACI uniquely combines fixed-time convergence with H_2/H_∞ optimization and concurrent learning, providing superior guarantees for both nominal operation and under uncertainties.

Algorithm 1 Fixed-time Concurrent Learning-based Actor-Critic-Identifier Control

Initialize:
  ACI neural networks:
    - Critic/Actor weights $\hat{W}_{ci}, \hat{W}_{ai}$ $(i = 1, 2)$
    - Identifier parameters $\hat{W}_\theta$
  Learning parameters:
    - Learning rates $k_{ci,j}, k_{ai,j} (i = 1, 2, j = 1, 2)$
    - Gain matrices $\Gamma_\theta, \Gamma_{ai}, \Gamma_{ci} (i = 1, 2)$
    - Fractional exponents γ_1, γ_2
    - Buffer sizes N
  Simulation time T_{end}
Online Learning:
while $t < T_{end}$ **do**
  // State Measurement & Reference
    Obtain current state X and reference X_d
  // Policy Approximation
    Compute control input $\hat{U}(X)$ via (26)
    Compute disturbance $\hat{\omega}(X)$ via (27)
  // System Identification
    Estimate dynamics $\hat{F}(X)$ via (29)
    Compute ID error ε_θ via (30)
  // Learning Error Computation
    Calculate Bellman errors:
      - Leader: $\hat{\varepsilon}_1$ via (32)
      - Follower: $\hat{\varepsilon}_2$ via (33)
  // Experience Replay Update
    Update buffers:
      $\mathscr{D}_1(t) \leftarrow \{\hat{U}, \hat{\varepsilon}_1, \{\hat{U}^j, \hat{\varepsilon}_1^j\}_{j=1}^N\}$
      $\mathscr{D}_2(t) \leftarrow \{\hat{\omega}, \hat{\varepsilon}_2, \{\hat{\omega}^j, \hat{\varepsilon}_2^j\}_{j=1}^N\}$
      $\mathscr{D}_\theta(t) \leftarrow \{\hat{F}, \varepsilon_\theta, \{\hat{F}^j, \varepsilon_\theta^j\}_{j=1}^N\}$
  // Neural Network Updates
    Update weights:
      - Critics: $\hat{W}_{ci}$ via (35)
      - Actors: $\hat{W}_{ai}$ via (36)
      - Identifier: $\hat{W}_\theta$ via (31)
  // Control Execution
    Apply control input $\hat{U}(X)$ to system
end while

Remark 8 (Handling Sequential Policy Updates) Sequential policy updates in Stackelberg games may cause oscillations, slow convergence, and jerky control signals due to players reacting to outdated information. Our FxT-CL-ACI framework addresses these challenges through: (1) Fractional power signum functions for smooth convergence behavior; (2) Two-timescale learning with faster controller updates than disturbance

Journal: 11071 MS: 11235 TYPESET DISK LE CP Disp.:2025/5/2 Pages: 21 Layout: Medium

Table 2 Comparison of FxT-CL-ACI with existing methods

Method	FxT Converge	Concurrent Learning	H_∞ Optimizat	Uncertainty Handling	Input Constraints
H_∞-ADP [22,23,50]	✗	✓	✓	✗	✓
CL-SysID [4,36]	✓	✓	✗	✓	✗
Classical CL [51]	✗	✓	✗	✗	✗
Regular ADP [24,52]	✗	✗	✗	✗	✓
Our FxT-CL-ACI	✓	✓	✓	✓	✓

model updates; (3) Concurrent learning with prioritized experience replay to reduce outdated sample influence; and (4) Constrained neural network weights to prevent control signal jumps. These mechanisms ensure stable performance despite the sequential nature of game-theoretic optimization.

Remark 9 (Identification vs. Integral RL) While integral reinforcement learning (IRL) could eliminate explicit F(x) identification as in [22,50], our identifier-based approach offers key advantages for UAV applications. First, it provides fixed-time rather than asymptotic convergence guarantees. Second, it delivers superior robustness against matched and unmatched uncertainties. Third, it offers better computational efficiency by avoiding complex activation function integrals. To our knowledge, fixed-time IRL remains unexplored in literature. Our future work aims to develop such a framework [22,23], combining IRL's model-free advantages with fixed-time convergence guarantees.

4.4 Stability analysis

In this subsection, we prove that under the proposed robust optimal tracking control scheme, the closed-loop system states and actor-critic NN estimation errors are ultimately uniformly bounded (UUB). We first introduce three key assumptions required for the stability analysis.

Definition 4 (Ultimate Uniform Boundedness [14,58]) A solution $x(t)$ of a system is said to be ultimately uniformly bounded (UUB) if there exist positive constants b and c, independent of initial time $t_0 \geq 0$, and for any $a \in (0, c)$, there exists a positive constant $T = T(a, b)$, such that $\|x(t_0)\| \leq a$ implies $\|x(t)\| \leq b$ for all $t \geq t_0 + T$.

Assumption 4.1 (Neural Network Boundedness [31, 32]) The neural network parameters satisfy the following uniform boundedness conditions ($i = 1, 2$):

1. The critic networks satisfy:

$$\|\hat{W}_{ci}\| \leq W_{Hci}, \|\varphi_{ci}(X)\| \leq \varphi_{Hci},$$
$$\|\nabla\varphi_{ci}(X)\| \leq \varphi_{D,Hci}, \|\delta_{ci}(X)\| \leq \delta_{Hci},$$
$$\|\nabla\delta_{ci}(X)\| \leq \delta_{D,Hci}$$

2. The actor networks satisfy:

$$\|\hat{W}_{ai}\| \leq W_{Hai}, \|\varphi_{ai}(X)\| \leq \varphi_{Hai},$$
$$\|\nabla\varphi_{ai}(X)\| \leq \varphi_{D,Hai}, \|\delta_{ai}(X)\| \leq \delta_{Hai},$$
$$\|\nabla\delta_{ai}(X)\| \leq \delta_{D,Hai}$$

3. The identifier networks satisfy:

$$\|\hat{W}_{\theta}\| \leq W_{H\theta}, \|\varphi_{\theta}(X)\| \leq \varphi_{H\theta},$$
$$\|\nabla\varphi_{\theta}(X)\| \leq \varphi_{D,H\theta}, \|\delta_{\theta}(X)\| \leq \delta_{H\theta},$$
$$\|\nabla\delta_{\theta}(X)\| \leq \delta_{D,H\theta}$$

where $W_{H*}, \varphi_{H*}, \varphi_{D,H*}, \delta_{H*}, \delta_{D,H*}, \sigma_{H*}, \sigma_{D,H*}$ are positive constants, and the upper bound of approximation errors $\delta_{D,H} = \max\{\delta_{D,Hc1}, \delta_{D,Hc2}, \delta_{D,Ha1}, \delta_{D,Ha2}, \delta_{D,H\theta}\}$.

Assumption 4.2 (Persistent Excitation [52,59]) For each agent $i = 1, 2$, the online and historical data must satisfy the following excitation conditions to ensure sufficient learning: (1) Online Data Excitation:

$$\Lambda_{1,i}\mathscr{I}_{m,i} \leqslant \int_t^{t+T} \rho_i(\tau)\sigma_i(\tau)^\top d\tau, \ \forall t \geq t_0, i = 1, 2 \tag{37}$$

(2) Historical Data Excitation:

$$\Lambda_{2,i}\mathscr{I}_{m,i} \leqslant \inf_{t \geq t_0} \frac{1}{N}\sum_{k=1}^{N} \rho_i^k(t)\sigma_i^k(t)^\top, \ \forall t \geq t_0, i = 1, 2 \tag{38}$$

Journal: 11071 MS: 11235 TYPESET DISK LE CP Disp.:2025/5/2 Pages: 21 Layout: Medium

(3) Combined Data Excitation:

$$\Lambda_{3,i}\mathscr{I}_{m,i} \leqslant \sum_{k=1}^{N}\int_{t}^{t+T}\frac{\rho_i^k(\tau)\sigma_i^k(\tau)^\top}{N}d\tau,\ \forall t \ge t_0, i=1,2 \tag{39}$$

where $\mathscr{I}_{m,i}$ is the identity matrix of dimension m, and at least one excitation measure $\Lambda_{j,i}$ $(j=1,2,3)$ must be strictly positive to guarantee sufficient exploration for learning convergence.

Based on the controller (26) and disturbance (27) designs, we have:

$$\|U^*(X)-\hat{U}(X)\|^2 \le \Sigma_1\tilde{W}_{a1}^\top\tilde{W}_{a1}+\Pi_1 \tag{40}$$

$$\|\omega^*(X)-\hat{\omega}(X)\|^2 \le \Sigma_2\tilde{W}_{a2}^\top\tilde{W}_{a2}+\Pi_2 \tag{41}$$

where Σ_i depends on upper bouds $\varphi_{H,i}$, $\varphi_{D,Hi}$, σ_{Hi}, $\sigma_{D,Hi}$, and Π_i depends on upper boud $\delta_{D,Hi}$.

Remark 10 (Derivation of Control Policy Error Bounds) The upper bounds in equations (40)-(41) representing the squared error between optimal policies ($U^*(X)$, $\omega^*(X)$) and their estimates ($\hat{U}(X)$, $\hat{\omega}(X)$) are derived as follows:

$$\begin{aligned}\|U^*(X)-\hat{U}(X)\|^2 &= \|-\frac{1}{2}R_1^{-1}G^\top(\tilde{W}_{a1}^\top\varphi_{a1}+\delta_{a1})\|^2\\ &\le \frac{1}{4}\|R_1^{-1}G^\top\|^2\cdot\|\varphi_{a1}\|^2\cdot\|\tilde{W}_{a1}\|^2\\ &\quad+\frac{1}{4}\|R_1^{-1}G^\top\|^2\cdot\|\delta_{a1}\|^2\\ &\le \frac{1}{4}\|R_1^{-1}G^\top\|^2\cdot\varphi_{H,1}^2\cdot\|\tilde{W}_{a1}\|^2\\ &\quad+\frac{1}{4}\|R_1^{-1}G^\top\|^2\cdot\delta_{D,H1}^2\\ &= \Sigma_1\tilde{W}_{a1}^\top\tilde{W}_{a1}+\Pi_1\end{aligned}$$

where $\Sigma_1 = \frac{1}{4}\|R_1^{-1}G^\top\|^2\cdot\varphi_{H,1}^2$ depends on the upper bounds of the activation functions $\varphi_{H,1}$, and $\Pi_1=\frac{1}{4}\|R_1^{-1}G^\top\|^2\cdot\delta_{D,H1}^2$ depends on the upper bound of the approximation error $\delta_{D,H1}$. Similarly, for the disturbance policy:

$$\|\omega^*(X)-\hat{\omega}(X)\|^2 \le \Sigma_2\tilde{W}_{a2}^\top\tilde{W}_{a2}+\Pi_2$$

where Σ_2 and Π_2 are analogously defined based on the bounds $\varphi_{H,2}$ and $\delta_{D,H2}$. These analytical bounds establish the relationship between neural network weight estimation errors and policy approximation errors, which is crucial for the stability analysis in Sect. 4.4.

The main stability result is given in the following theorem:

Remark 11 (Practical Verification of Assumptions) While Assumptions 4.1 and 4.2 provide theoretical guarantees for our approach, their practical verification is equally important: For Assumption 4.1, we employ neural networks with proper design, which is referred to literature [47,60], to ensure approximation capabilities within the compact set of interest. The approximation errors in our simulation and experimental results remain within the theoretically predicted bounds, validating this assumption. For Assumption 4.2, the Persistent Excitation (PE) condition is essential for concurrent learning stability. To ensure this condition is met, our implementation includes a large-size history stack, in which data samples N is selected as 30 lature [24,51] inspired by in both simulation and experimental setups. This ensures that the rank condition is satisfied throughout operation, as evidenced by the consistent convergence behavior observed in our experiments.

Theorem 4.3 *(Traditional Actor-Critic ($\gamma_1=0$, $\gamma_2=1$) [61]) Consider the closed-loop system (1) under the proposed FxT-CL-ACI control scheme in Algorithm 1. Let Assumptions 2.1-4.2 hold, $\gamma_1=0$ and $\gamma_2=1$, and the system parameters are known. Then the FxT-CL-ACI reduces to the traditional Actor-Critic (AC) control scheme.*

If the AC neural networks are updated according to (35) and (36), with control and disturbance policies estimated by (26) and (27), then the closed-loop system state X and all weight estimation errors remain ultimately uniformly bounded (UUB) if:

$$\|Z\| \ge \sqrt{\frac{\gamma_{\text{res}}}{\underline{\lambda}_{\mathscr{H}}}} \tag{42}$$

where:

- *$Z=[X^\top,\tilde{W}_{c1}^\top,\tilde{W}_{c2}^\top,\tilde{W}_{a1}^\top,\tilde{W}_{a2}^\top]^\top$ is the augmented error state vector containing tracking errors and weight estimation errors*
- *γ_{res} is the residual approximation error bound arising from neural network reconstruction errors defined as:*

$$\gamma_{\text{res}}=\frac{1}{2}k_{c1,1}\left(\frac{1}{4}\tilde{W}_{a1}^\top G_\sigma\tilde{W}_{a1}+\xi_{H1}+\Delta_1\right)^2+\gamma^2\Pi_{u_2}$$

$$
+\frac{1}{2}k_{c2,1}\left(\frac{1}{4}\tilde{W}_{a2}^\top K_\sigma \tilde{W}_{a2} - \frac{1}{4}\tilde{W}_{a1}^\top G_\sigma \tilde{W}_{a1} + \Delta_2\right)^2
+\frac{1}{2}k_{c1,2}\left(\frac{1}{4}\tilde{W}_{a1}^\top G_{\sigma,k}\tilde{W}_{a1} + \Delta_1^k\right)^2 + \bar{\lambda}_{R,1}\Pi_{u_1}
+\frac{1}{2}k_{c2,2}\left(\frac{1}{4}\tilde{W}_{a2}^\top K_{\sigma,k}\tilde{W}_{a2} - \frac{1}{4}\tilde{W}_{a1}^\top G_{\sigma,k}\tilde{W}_{a1} + \Delta_2^k\right)^2
$$

– $\underline{\lambda}_{\mathscr{H}}$ *is the minimum eigenvalue of matrix* $\mathscr{H}$ *defined as:*

$$
\mathscr{H} = \begin{bmatrix} h_1 & 0 & 0 & 0 & 0 \\ 0 & h_2 & 0 & 0 & 0 \\ 0 & h_3 & h_4 & 0 & 0 \\ 0 & h_5 & 0 & h_6 & 0 \\ 0 & 0 & h_7 & 0 & h_8 \end{bmatrix}
$$

with elements h_1 *to* h_8 *defined as:* $h_1 = \underline{\lambda}_{Q1} - \underline{\lambda}_{Q2}$, $h_2 = \frac{1}{2}k_{c1,1}\sigma_1\sigma_1^\top + \frac{1}{2}k_{c1,2}\Lambda_{2,1}\mathscr{I}_{m,1}$, $h_3 = (k_{c1,1} + k_{c2,1})\sigma_1\sigma_2^\top$, $h_4 = \frac{1}{2}k_{c2,1}\sigma_2\sigma_2^\top + \frac{1}{2}k_{c2,2}\Lambda_{2,2}\mathscr{I}_{m,2}$, $h_5 = -\Gamma_{a1}\mathscr{I}_{m,1}$, $h_6 = \Gamma_{a1}\mathscr{I}_{m,1} - \bar{\lambda}_{R,1}\Sigma_{u_1}\mathscr{I}_{m,1}$, $h_7 = -\Gamma_{a2}\mathscr{I}_{m,2}$, $h_8 = \Gamma_{a2}\mathscr{I}_{m,2} + \gamma^2\Sigma_{u_2}\mathscr{I}_{m,2}$.

Proof Consider the following Lyapunov function candidate for the time-varying closed-loop system:

$$
\mathscr{V}(X,t) = \sum_{i=1}^{2}\left(J_i^*(X,t) + \frac{1}{2}\tilde{W}_{ci}^\top(t)\tilde{W}_{ci}(t) + \frac{1}{2}\tilde{W}_{ai}^\top(t)\tilde{W}_{ai}(t)\right)
$$

where $J_i^*(X,t)$ is the optimal value function for agent i at time t. This Lyapunov function is positive definite and radially unbounded, satisfying $\mathscr{V}(0,t) = 0$ and $\mathscr{V}(X,t) > 0$ for all $X \neq 0$ and $t \geq 0$. Its time-varying nature accounts for neural network weight adaptation, changing references, and external disturbances in the closed-loop system. The approximated Bellman errors for both leader and follower agents can be expressed as:

$$
\begin{cases}
\hat{\varepsilon}_1 = -\sigma_1^\top \tilde{W}_{c1} + \frac{1}{4}\tilde{W}_{a1}G_\sigma\tilde{W}_{a1} + \Delta_1 + \xi_{H1} \\
\hat{\varepsilon}_2 = -\sigma_2^\top \tilde{W}_{c2} + \frac{1}{4}\left(\tilde{W}_{a2}K_\sigma\tilde{W}_{a2} - \tilde{W}_{a1}G_\sigma\tilde{W}_{a1}\right) \\
\qquad + \Delta_2 + \xi_{H2} \\
\hat{\varepsilon}_1^k = -(\sigma_1^k)^\top \tilde{W}_{c1} + \frac{1}{4}\tilde{W}_{a1}G_\sigma^k\tilde{W}_{a1} + \Delta_1^k \\
\hat{\varepsilon}_2^k = -(\sigma_2^k)^\top \tilde{W}_{c2} + \frac{1}{4}\left(\tilde{W}_{a2}K_\sigma^k\tilde{W}_{a2} - \tilde{W}_{a1}G_\sigma^k\tilde{W}_{a1}\right) + \Delta_2^k
\end{cases}
$$

where $G_\sigma = \nabla\varphi_{a1}^\top G R_1^{-1} G^\top \nabla\varphi_{a1}^\top$, $K_\sigma = \frac{1}{2\gamma^2}\nabla\varphi_{a2}^\top K R_2^{-1} K^\top \nabla\varphi_{a1}^\top$, $G_\sigma^k = G_\sigma(X^k)$, $K_\sigma^k = K_\sigma(X^k)$, and Δ_i, Δ_i^k represent uniformly bounded approximation errors. Taking the time derivative of $\mathscr{V}$ along system trajectories:

$$
\dot{\mathscr{V}} = \sum_{i=1}^{2}\left[\nabla J_i^*\left(F + GU + K\omega\right) + \tilde{W}_{ci}^\top \dot{\tilde{W}}_{ci}^\top + \tilde{W}_{ai}^\top \dot{\tilde{W}}_{ai}^\top\right]
$$

Substituting the weight update laws from (35) and (36):

$$
\dot{\mathscr{V}} = \sum_{i=1}^{2}\Big[\nabla J_i^*\left(F + GU + K\omega\right)
- \tilde{W}_{ci}^\top k_{ci,1}\rho_i[\hat{\varepsilon}_i + \hat{\varepsilon}_i] - \tilde{W}_{ci}^\top \frac{k_{ci,2}}{N}\sum_{k=1}^{N}\rho_i^k[\hat{\varepsilon}_i^k
+ \hat{\varepsilon}_i^k] - \tilde{W}_{ai}^\top \Gamma_{ai}[k_{ai,1}\varepsilon_{ai} + k_{ai,2}\varepsilon_{ai}]\Big]
$$

Substituting the Hamiltonian functions and Bellman errors yields:

$$
\begin{aligned}
\dot{\mathscr{V}} \leq & -X^\top\left(Q_1 - Q_2\right)X - \Phi(U) - \gamma^2\|\omega\|^2 - \eta^\top\dot{\lambda}_2 \\
& - \sum_{i=1}^{2}k_{ci,1}\tilde{W}_{ci}^\top\frac{\sigma_i}{\rho_i}\left(-\sigma_i^\top\tilde{W}_{ci} + \frac{1}{4}\tilde{W}_{ai}^\top\Sigma_i\tilde{W}_{ai} + \Delta_i\right) \\
& - \sum_{i=1}^{2}k_{ai,1}\tilde{W}_{ai}^\top\Gamma_{ai}\left(\hat{W}_{ai} - \hat{W}_{ci}\right) \\
& - \sum_{i=1}^{2}\frac{k_{ci,2}}{N}\tilde{W}_{ci}^\top\sum_{k=1}^{N}\frac{\sigma_i^k}{\rho_i^k}\left(-(\sigma_i^k)^\top\tilde{W}_{ci} + \Delta_i^k\right) \qquad (43)
\end{aligned}
$$

Leveraging the PE conditions from Assumption 4.2, these PE conditions ensure sufficient richness in both current and historical data samples, guaranteeing that:

$$
\tilde{W}_{ci}^\top k_{ci,1}\rho_i\sigma_i^\top\tilde{W}_{ci} \geq k_{ci,1}\Lambda_{1,i}\|\tilde{W}_{ci}\|^2 \qquad (44)
$$

$$
\tilde{W}_{ci}^\top\frac{k_{ci,2}}{N}\sum_{k=1}^{N}\rho_i^k(\sigma_i^k)^\top\tilde{W}_{ci} \geq k_{ci,2}\Lambda_{2,i}\|\tilde{W}_{ci}\|^2 \qquad (45)
$$

Then using (40), (41) and Young's inequality, we obtain:

$$
\begin{aligned}
\dot{V}(Z) &= Z^T\mathscr{H}Z + \gamma_{\text{res}} \\
&\leq -\underline{\lambda}_{\mathscr{H}}\|Z\|^2 + \gamma_{\text{res}}
\end{aligned}
$$

When $\|Z\| > \sqrt{\frac{\gamma_{\text{res}}}{\underline{\lambda}_{\mathscr{H}}}}$, we have $\dot{V}(Z) < 0$, which forces the trajectory to enter and remain within the bounded region, satisfying the UUB definition. Therefore, when

Journal: 11071 MS: 11235 TYPESET DISK LE CP Disp.:2025/5/2 Pages: 21 Layout: Medium

condition (42) is satisfied, the closed-loop system state X and actor-critic estimation errors are ultimately uniformly bounded. □

Theorem 4.3 establishes the UUB property of the closed-loop system states and actor-critic NN estimation errors under the traditional AC scheme. The proof demonstrates that the proposed FxT-CL-ACI control scheme guarantees robust optimal tracking performance for the leader-follower agents in the Stackelberg game framework. Next, we extend the analysis to the fixed-time convergence case, where the learning errors of the ACI NNs converge to a bounded region in fixed time.

Theorem 4.4 *(FxT-CL-ACI ($0 < \gamma_1 < 1,\ \gamma_2 > 1$)) Consider the concurrent learning update law (31), (35) and (36) under the proposed FxT-CL-ACI control scheme in Algorithm 1. Let Assumptions 2.1-4.2 hold, and suppose the following condition is satisfied:*

$$\sqrt{\frac{2\bar{\lambda}_\Gamma}{(2\underline{\lambda}_\Gamma)^{\gamma_2+1}}} < \frac{\alpha_1}{\beta} \tag{46}$$

where $\Gamma = diag([\Gamma_{c1}, \Gamma_{c2}, \Gamma_{a1}, \Gamma_{a2}, \Gamma_\theta])$ is the gain matrix. Then the estimation errors of the actor-critic-identifier NN weights $\tilde{W} = [\tilde{W}_{c1}^\top, \tilde{W}_{c2}^\top, \tilde{W}_{a1}^\top, \tilde{W}_{a2}^\top, \tilde{W}_\theta^\top]^\top$ converge to a bounded region in fixed time:

$$\|\tilde{W}(t)\| \le \sqrt{\frac{\bar{\lambda}_\Gamma}{\underline{\lambda}_\Gamma}} \min\{\sqrt{2\bar{\lambda}_\Gamma}, \bar{\xi}\} \quad \forall t \ge T \tag{47}$$

where the convergence time T is bounded by $T_{\max}$:

$$T_{\max} = \frac{2}{\alpha(\gamma_2 - 1)} + \frac{2(1 - (2\underline{\lambda}_\Gamma)^{-1/2} \min\{\bar{\xi}, \sqrt{2\bar{\lambda}_\Gamma}\})^{1-\gamma_1})}{\alpha_2(1-\delta)(1-\gamma_1)(2\underline{\lambda}_\Gamma)^{(\gamma_1+1)/2}} \tag{48}$$

with $\bar{\xi}$ defined as:

$$\bar{\xi} = \max\left\{\frac{\delta_{D,H}}{\min\{\underline{\lambda}_{\Psi(t)}^{1/2}, \bar{\lambda}_h\}}, \left(\frac{\beta}{\alpha_2\delta}\right)^{1/\gamma_1}\right\} \tag{49}$$

and the coefficients defined as:

$$\alpha = \alpha_1(2\underline{\lambda}_\Gamma)^{(\gamma_2+1)/2} - \beta\sqrt{2\bar{\lambda}_\Gamma} \tag{50}$$

$$\alpha_1 = 2^{1-\gamma_2} n^{(1-\gamma_2)/2}(K_1 \Lambda_{1,i}^{(\gamma_2+1)/2} + K_2 \Lambda_{2,i}^{(\gamma_2+1)/2}) \tag{51}$$

$$\alpha_2 = K_1 \Lambda_{1,i}^{(\gamma_1+1)/2} + K_2 \Lambda_{2,i}^{(\gamma_1+1)/2} + K_3 \Lambda_{3,i}^{(\gamma_1+1)/2} \tag{52}$$

$$\beta = (K_1 + K_2 + K_3)[n^{(2-\gamma_1)/4}\delta_{D,H}^{\gamma_1} + \delta_{D,H}^{\gamma_2}] \tag{53}$$

with additional terms defined as $\varphi = [\varphi_{c1}^\top, \varphi_{c2}^\top, \varphi_{a1}^\top, \varphi_{a2}^\top, \varphi_\theta^\top]^\top$, $\delta = [\delta_{c1}^\top, \delta_{c2}^\top, \delta_{a1}^\top, \delta_{a2}^\top, \delta_\theta^\top]^\top$, $\Psi(t) = \varphi(t)\varphi^\top(t)$ corresponds to instantaneous data excitation with lower bound $\Lambda_{1,i}$, $\Theta = \frac{1}{N}\sum_{k=1}^N \varphi(\tau_k)\varphi^\top(\tau_k)$ represents historical data excitation with lower bound $\Lambda_{2,i}$.

Proof Consider the Lyapunov function candidate:

$$\mathscr{V}(t) = \frac{1}{2}\tilde{W}^\top(t)\Gamma^{-1}\tilde{W}(t) \tag{54}$$

Taking the time derivative of $\mathscr{V}(t)$, the following expression is obtained:

$$\begin{aligned}\dot{\mathscr{V}}(t) = \mathrm{tr}\{&-K_1\tilde{W}^\top(t)\varphi(t)(\lfloor \varphi^\top(t)\tilde{W}(t) - \delta^\top(t)\rceil^{\gamma_1} \\ &+ \lfloor \varphi^\top(t)\tilde{W}(t) - \delta^\top(t)\rceil^{\gamma_2}) \\ &- \frac{K_2}{N}\tilde{W}^\top(t)\sum_{k=1}^N \varphi(\tau_k)(\lfloor \varphi^\top(\tau_k)\tilde{W}(t) - \delta^\top(\tau_k)\rceil^{\gamma_1} \\ &+ \lfloor \varphi^\top(\tau_k)\tilde{W}(t) - \delta^\top(\tau_k)\rceil^{\gamma_2})\}\end{aligned} \tag{55}$$

where $K_1 = \mathrm{diag}([k_{ci,1}, k_{ai,1}, k_\theta])$, $K_2 = \mathrm{diag}([k_{ci,2}, k_{ai,2}, k_\theta])$. For the case where $|(\varphi^\top\tilde{W})_i| \ge |\delta_i|$, we have $\mathrm{sign}(\varphi^\top\tilde{W} - \delta^\top) = \mathrm{sign}(\varphi^\top\tilde{W})$. For $0 \le \gamma_1 < 1$, using the inequality $|y_1 + y_2|^{\gamma_1} \le |y_1|^{\gamma_1} + |y_2|^{\gamma_1}$, we obtain:

$$\begin{aligned}&|(\varphi^\top(t)\tilde{W}(t))_i|^{\gamma_1} \\ &\quad \le |(\varphi^\top(t)\tilde{W}(t))_i - \delta_i(t)|^{\gamma_1} + |\delta_i(t)|^{\gamma_1}\end{aligned}$$

Also, for $\gamma_2 > 1$, using $|y_1 + y_2|^{\gamma_2} \le 2^{\gamma_2-1}(|y_1|^{\gamma_2} + |y_2|^{\gamma_2})$, we have:

$$\begin{aligned}&-|\varphi^\top(t)\tilde{W}(t) - \delta(t)|^{\gamma_2} \\ &\quad \le -2^{1-\gamma_2}|\varphi^\top(t)\tilde{W}(t)|^{\gamma_2} + |\delta(t)|^{\gamma_2}\end{aligned}$$

Using the PE conditions from Assumption 4.2, the time derivative of $\mathscr{V}(t)$ can be further bounded as:

$$\dot{\mathscr{V}}(t) \le -\alpha_2\|\tilde{W}(t)\|^{\gamma_1+1} - \alpha_1\|\tilde{W}(t)\|^{\gamma_2+1} + \beta\|\tilde{W}(t)\| \tag{56}$$

Journal: 11071 MS: 11235 TYPESET DISK LE CP Disp.:2025/5/2 Pages: 21 Layout: Medium

where the coefficients now explicitly incorporate the PE measures:

$$\alpha_1 = 2^{1-\gamma_2} n^{(1-\gamma_2)/2} (K_1 \Lambda_{1,i}^{(\gamma_2+1)/2} + K_2 \Lambda_{2,i}^{(\gamma_2+1)/2}) \tag{57}$$

$$\alpha_2 = K_1 \Lambda_{1,i}^{(\gamma_1+1)/2} + K_2 \Lambda_{2,i}^{(\gamma_1+1)/2} + K_3 \Lambda_{3,i}^{(\gamma_1+1)/2} \tag{58}$$

$$\beta = (K_1 + K_2 + K_3)[n^{(2-\gamma_1)/4} \delta_{D,H}^{\gamma_1} + \delta_{D,H}^{\gamma_2}] \tag{59}$$

Note that $\Psi(t) = \varphi(t)\varphi^\top(t)$ corresponds to instantaneous excitation with lower bound $\Lambda_{1,i}$, $\Theta = \frac{1}{N}\sum_{k=1}^{N} \varphi(\tau_k)\varphi^\top(\tau_k)$ represents historical data excitation with lower bound $\Lambda_{2,i}$, and the integrated excitation over time interval $[t, t+T]$ has lower bound $\Lambda_{3,i}$. Then, for $\mathscr{V}(t) > 1$, we have the following inequality holding:

$$\dot{\mathscr{V}}(t) \le -\alpha \mathscr{V}^{(\gamma_2+1)/2}(t) \tag{60}$$

where $\alpha = \alpha_1 (2\underline{\lambda}_\Gamma)^{(\gamma_2+1)/2} - \beta\sqrt{2\bar{\lambda}_\Gamma}$ is positive when:

$$\frac{\alpha_1}{\beta} > \sqrt{\frac{2\bar{\lambda}_\Gamma}{(2\underline{\lambda}_\Gamma)^{\gamma_2+1}}} \tag{61}$$

Then, the Lyapunov function $\mathscr{V}(t)$ converges to a bounded region in fixed time $T \le T_{\max}$, where the states are fixed-time attractive with bound:

$$\|\tilde{W}(t)\| \le \sqrt{\bar{\lambda}_\Gamma / \underline{\lambda}_\Gamma} \min\{\sqrt{2\bar{\lambda}_\Gamma}, \bar{\xi}\} \tag{62}$$

where:

$$\bar{\xi} = \max\left\{ \frac{\delta_{D,H}}{\min\{\underline{\lambda}_{\Psi(t)}^{1/2}, \bar{\lambda}_h\}}, (\frac{\omega}{\alpha_2 \delta})^{1/\gamma_1} \right\} \tag{63}$$

This completes the proof showing fixed-time convergence of proposed FxT-CL-ACI control scheme's learning process. □

5 Simulations verification

In this section, we validate the effectiveness of the proposed FxT-CL-ACI control scheme through comprehensive numerical simulations.

5.1 Simulation setup

Consider an uncertain nonlinear system with drift dynamics (1) in the form:

$$f = \begin{bmatrix} x_1 & x_2 & 0 & 0 \\ 0 & 0 & x_1 & x_2(\cos(2x_1)+2) \end{bmatrix} \times \begin{bmatrix} W_\theta(1) \\ W_\theta(2) \\ W_\theta(3) \\ W_\theta(4) \end{bmatrix} \tag{64}$$

$$g = \begin{bmatrix} \sin(2x_1+1)+2 & 0 \\ 0 & \cos(2x_1)+2 \end{bmatrix}, k = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \tag{65}$$

where the actual value of the unknown drift parameters are set as $W_\theta = [-1, 1, -0.5, -0.5]^\top$, the basis function of the identifier is defined as $\varphi_\theta = [X_1, X_2, X_1, x2(\cos(2x_1)+2)]^\top$. The approximate optimal control input is computed from (26), and the approximate worst disturbance is derived from (27). The actor-critic neural networks are updated according to (35) and (36). The detailed NN setup is as follows:

- Basis functions φ_{ij} $(i = c, a, j = 1, 2)$ are defined as:

$$\varphi_{ij} = \frac{1}{\alpha+1}\Big[X_1^{\alpha+1}, (X_1X_2)^{\alpha+1}, X_2^{\alpha+1}, (X_1^2X_2)^{\alpha+1}, (X_1X_2^2)^{\alpha+1}, (X_1^2X_2^2)^{\alpha+1}\Big]^\top$$

  where X_i denotes the i-th state variable of the system, $\alpha \in (0, 1]$ is the fractional power. We choose $\alpha = 1$ in the simulation.
- Initial AC NN weights: $\hat{W}_{cij}(0) = \hat{W}_{aij}(0) = 1 (i = 1, 2, j = 1, \cdots, 6)$.

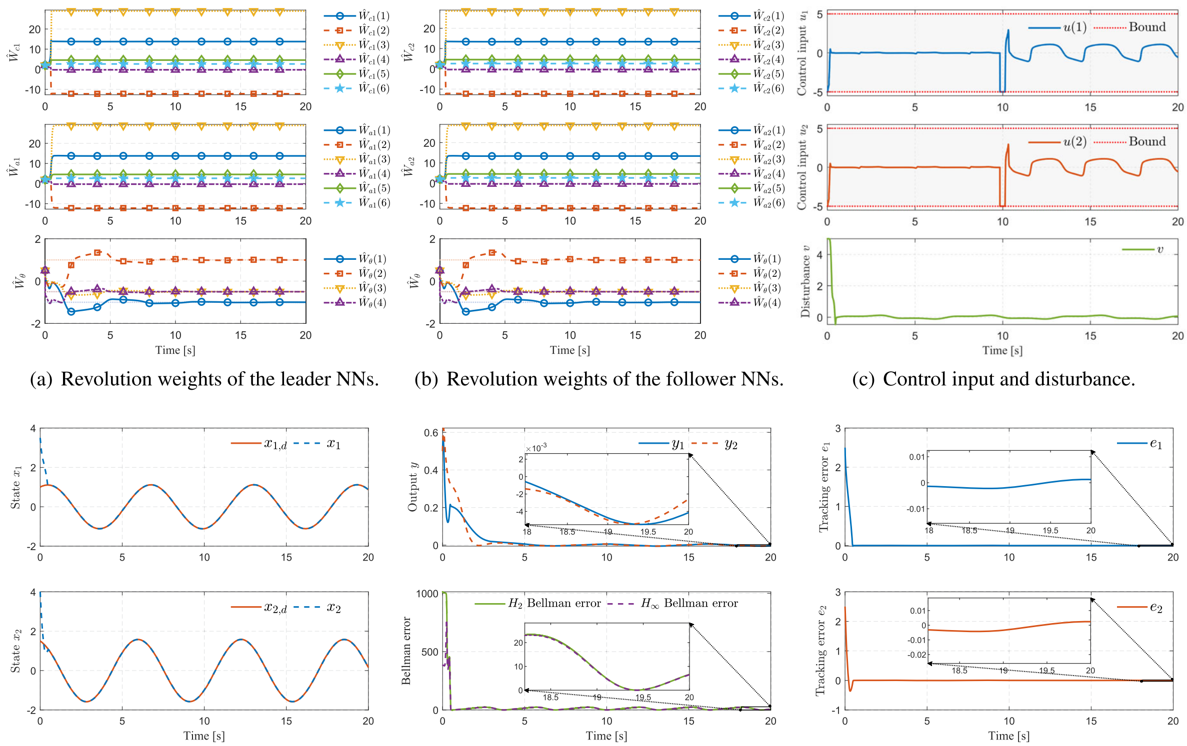
The complete set of control parameters is provided in Table 3. The simulation is conducted in MATLAB R2023b on a computer with Intel i3-12100 CPU and 24GB RAM. The simulation time is set to $T = 20$ seconds, and the ODE solver is set to the 4th-order Runge–Kutta method with a fixed time step of 0.001 seconds, while the random seed is set to 1 for reproducibility.

5.2 Simulation results

The simulation results demonstrating the effectiveness of the proposed FxT-CL-ACI scheme are shown in Fig. 2. The evolution of actor-critic NN weights is presented in Figs. 2a -2b, which show convergence of the

Table 3 Parameters of simulation cases

Component	Parameter	Value
Leader	Cost Matrix	$R_1 = \mathscr{I}_2,\quad Q_1 = 2\mathscr{I}_2$
	Critic Params	$k_{1,c1} = 0.5, k_{1,c2} = 0.1$
	Actor Params	$k_{1,ai} = 1,\quad \Gamma_j = \mathscr{I}_6$
Follower	Cost Matrix	$R_2 = \mathscr{I}_2,\quad Q_2 = 2\mathscr{I}_2$
	Critic Params	$k_{2,c1} = 0.5, k_{2,c2} = 0.1$
	Actor Params	$k_{2,ai} = 1,\quad \Gamma_j = \mathscr{I}_6$
Common	Fixed-time	$\gamma_1 = 0.8, \gamma_2 = 1.2$
	Other	$\gamma = 2,\quad \mu = 0.5$



(a) Revolution weights of the leader NNs. (b) Revolution weights of the follower NNs. (c) Control input and disturbance.

(d) The tracking performance of the system. (e) Value of function y and the Bellman error. (f) Tracking errors of the system.

Fig. 2 Results of the FxT-CL-ACI scheme in tracking control simulation

learning process. Figure 2c displays the control inputs and disturbances acting on the system. The tracking performance is illustrated in Fig. 2a, which shows that the system states successfully track their desired trajectories. he Bellman errors and costate function evolution are shown in Fig. 2b, validating the optimality of the learned control policy. Figure 2c presents the tracking errors, demonstrating that they remain bounded and converge to a small neighborhood of zero under the proposed control scheme.

6 Hardware experiments

In this section, a UAV-based physical experiments are conducted to further verify the effectiveness of the proposed FxT-CL-ACI control scheme.

Journal: 11071 MS: 11235 TYPESET DISK LE CP Disp.:2025/5/2 Pages: 21 Layout: Medium

Fig. 3 Hardware equipment used in the UAV tracking control experiment

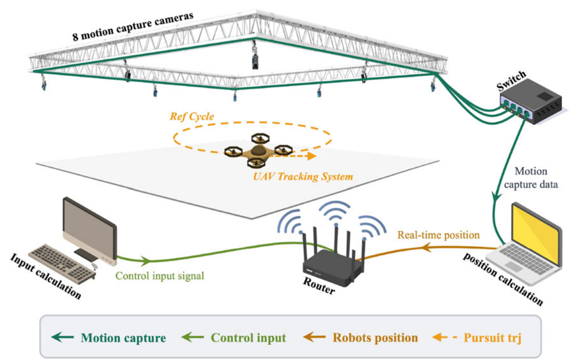


Fig. 4 Information flow in the UAV tracking system

6.1 Experimental setup

The experiment is conducted on an X150 quadrotor UAV platform to validate the trajectory tracking capabilities of the proposed control scheme. The integrated hardware system consists of the following components:

I. UAV Platform: The quadrotor UAV is equipped with an RK3566 quad-core processor and 4GB RAM for real-time computation. Four high-performance brushless DC motors with precision ESCs provide reliable actuation. A 9-axis IMU enables high-accuracy attitude estimation. A 5GHz dual-band WiFi module ensures reliable communication with the ground station. The dynamic model of the UAV could be formulated as $\dot{x} = V_x = u_x$, $\dot{y} = V_y = u_y$, and $u = [u_x, u_y]^\top$ is the control input vector in 2-dimensional space. The disturbance acting on the UAV is from the external wind and sensor noise. The UAV and ground workstation communicate state information through a 5GHz wireless network with a standard UDP protocol as illustrated in Fig. 4

II. Testing Environment: The experimental setup utilizes a professional OptiTrack motion capture system with 8 high-speed cameras providing sub-millimeter precision 6-DOF pose tracking at 120Hz. A dedicated ground control station (Intel i7-12700, 32GB RAM) runs the optimized motion capture software for real-time trajectory recording and controller implementation.

III. Control Implementation: The control system operates at 30Hz with deterministic timing ($\Delta t = 1/30s$). Velocity commands are transmitted via robust WiFi communication. High-rate state feedback is provided by the motion capture system at 120Hz. Online learning is efficiently executed on the ground station computer.

To enhance computational efficiency and learning convergence, we adopt an fractional-order finite-time neural network architecture for the actor-critic networks as:

$$\varphi_{ij} = \left[\frac{1}{\alpha+1} X_1^{\alpha+1}, \frac{1}{\alpha+1} X_2^{\alpha+1}, \frac{1}{\alpha+1} (X_1 X_2)^{\alpha+1} \right]^\top$$

which is proposed in [47,60]. All the initial network weights are configured as $\hat{W}_{ij} = 10$, and the other parameters are the same as in the simulation setup in Table 3. Figure 3 shows the complete hardware setup used in the experiments. This integrated system enables comprehensive validation of the proposed FxT-CL-ACI scheme under real-world conditions. To evaluate controller performance under realistic disturbances including wind effects, aerodynamic forces, and sensor noise, we design a circular reference trajectory with:

$$\begin{cases} \textit{Radius:} \quad r = 1.5 \text{ meters} \\ \textit{Period:} \quad T = 10\pi \approx 31.4 \text{ seconds} \end{cases} \tag{66}$$

This trajectory allows thorough assessment of the tracking capabilities and disturbance rejection properties of the proposed control scheme. The experimental results are shown in Fig. 5-7. Figure 5 shows the sketch of the UAV tracking the reference trajectory with high precision. The 3D trajectory tracking performance illustrated in Fig. 6 demonstrates accurate reference following capability.

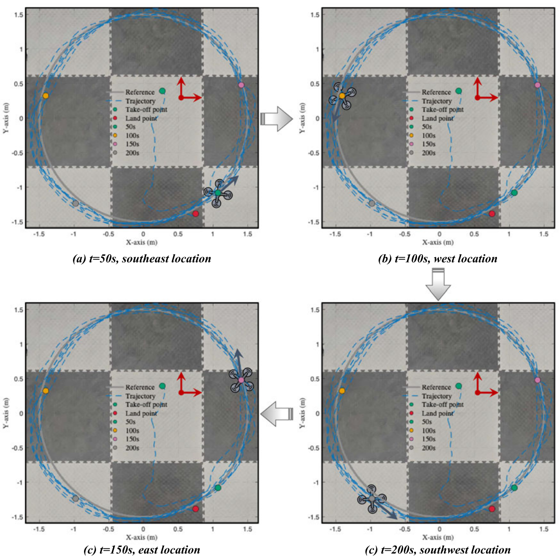


Fig. 5 Sketch of the UAV tracking the reference trajectory in the experiment (with sketch UAV representing position)

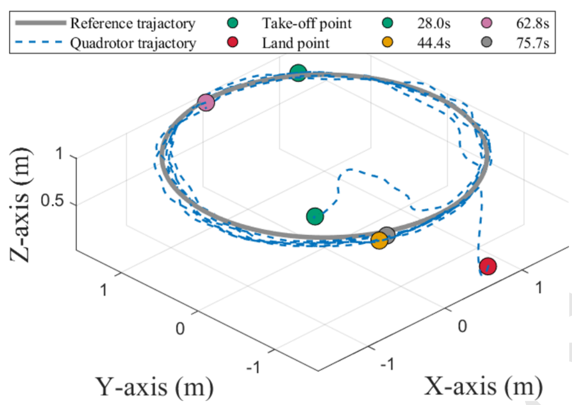


Fig. 6 Trajectory of the UAV in 3-dimensional space

6.2 Performance analysis

Basic Tracking Performance: As shown in Fig. 7a, the UAV achieves precise 2D trajectory tracking with an small error of ± 0.5 m. The tracking errors e_X and e_Y in Fig. 7b remain bounded within ± 0.6 m despite external disturbances. The evolution of critic NN weights in Fig. 7c shows rapid convergence within 100 s, validating the fixed-time learning property.

Control System Analysis: Fig. 7a shows that the control inputs remain within saturation bounds while achieving desired tracking performance. The UAV's Euler angles depicted in Fig. 7b exhibit smooth transitions during trajectory following. The 3D error distribution visualization in Fig. 7c reveals that most tracking errors are concentrated within a small region around the reference trajectory.

Advanced Performance Metrics: Fig. 7a analyzes the correlation between tracking velocity and position error, showing that higher velocities generally correspond to larger tracking errors. The statistical distribution of error peaks in Fig. 7b follows a correlation coefficient of $R = -0.024$, with mean velocity $v = 0.31$ m/s and error standard deviation $\sigma = 0.5681$ m. The energy consumption analysis in Fig. 7c demonstrates efficient performance with maximum kinetic energy of 0.5 J and power consumption of 0.34 W.

The experimental results of the UAV tracking system validate the robustness and energy efficiency of the proposed FxT-CL-ACI control scheme under real-world disturbances:

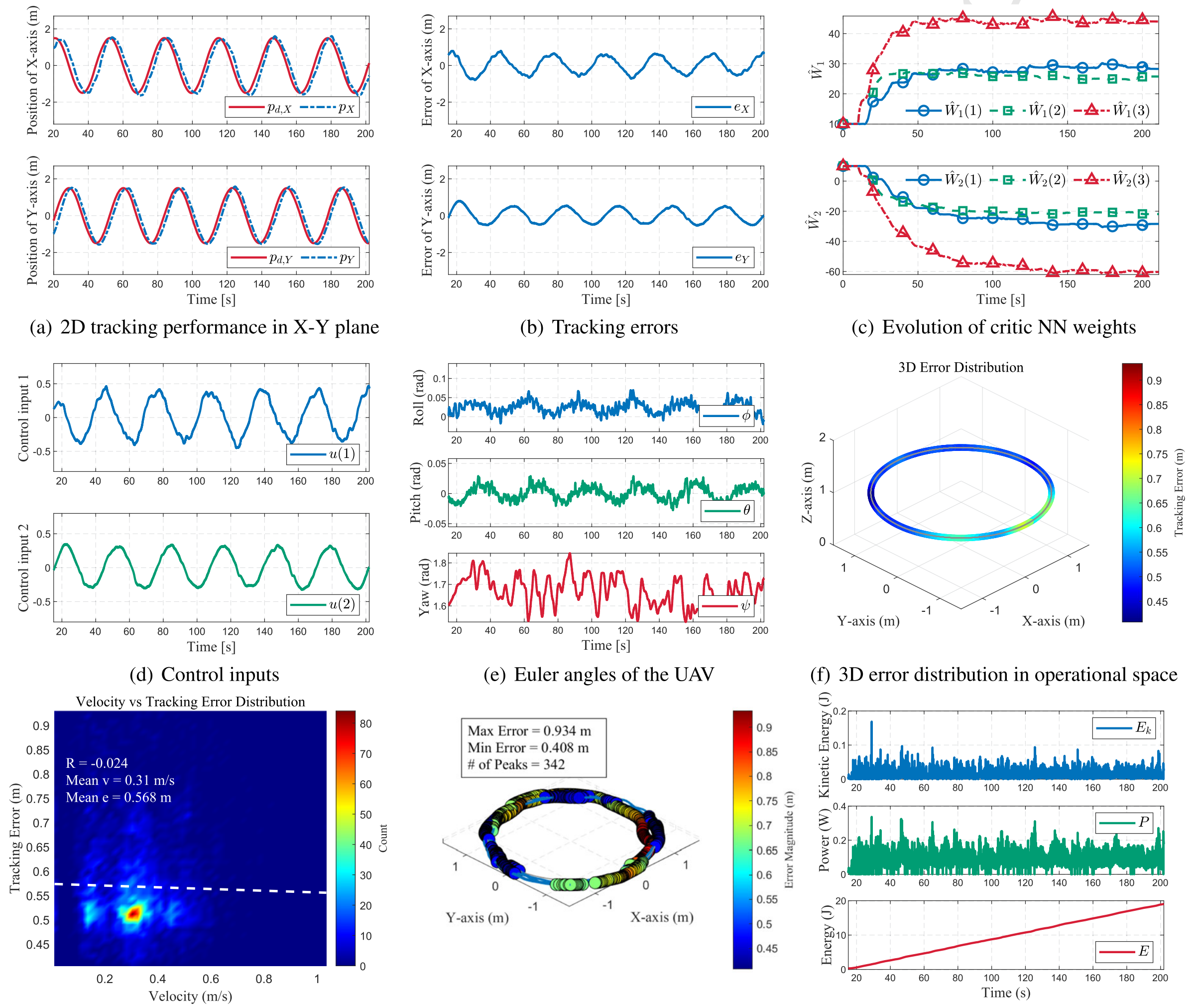
1. The proposed FxT-CL-ACI scheme achieves robust trajectory tracking with bounded errors under real-world disturbances
2. Fixed-time learning enables rapid convergence of neural network weights within 100 s
3. The control strategy effectively balances tracking accuracy and energy efficiency
4. The Stackelberg game framework successfully handles the trade-off between optimal tracking and disturbance rejection

These comprehensive experimental results demonstrate the practical effectiveness of the proposed control scheme for real-world UAV applications requiring both robust performance and energy efficiency.

7 Conclusion

This paper presents a novel fixed-time concurrent learning-based actor-critic-identifier (FxT-CL-ACI) control scheme for robust optimal tracking control of nonlinear systems with uncertainties and disturbances. A Stackelberg game framework is established to design the robust optimal tracking controller by sequential optimization of H_2 and H_∞ performance indices, addressing both tracking performance and disturbance rejection. An ACI architecture with FxT-CL is developed to approximate the optimal control solution while identifying uncertain system parameters online. The FxT convergence property ensures rapid learning. Lyapunov stability analysis proves that under the proposed

Journal: 11071 MS: 11235 TYPESET DISK LE CP Disp.:2025/5/2 Pages: 21 Layout: Medium



(a) 2D tracking performance in X-Y plane (b) Tracking errors (c) Evolution of critic NN weights

(d) Control inputs (e) Euler angles of the UAV (f) 3D error distribution in operational space

(g) Correlation between velocity and error (h) Statistical distribution of tracking errors (i) Energy consumption analysis during tracking

Fig. 7 Comprehensive performance analysis of the UAV tracking system

scheme, both closed-loop system states and ACI estimation errors are ultimately uniformly bounded with FxT convergence. Comprehensive validation through numerical simulations and UAV hardware experiments demonstrates the tracking capabilities and disturbance rejection properties of the proposed control scheme. Four key limitations of the current approach encompass:

1. **Computational Complexity:** The FxT-CL-ACI scheme requires significant computational resources for real-time implementation, which may limit its applicability on resource-constrained platforms.
2. **Parameterization Requirements:** The approach relies on appropriate parameterization of system dynamics and neural network structures, requiring domain expertise for effective implementation.
3. **Initialization Sensitivity:** While the method guarantees fixed-time convergence, the performance can still be influenced by initial weight selection and learning parameter tuning.
4. **Disturbance Model Limitations:** Performance depends on the accuracy of disturbance modeling within the Stackelberg game framework and may degrade under unmodeled disturbance patterns.

Future research directions include extending the proposed framework to stochastic systems and multi-agent coordination problems, and exploring the development of a fixed-time integral reinforcement learning (FxT-

IRL) framework that combines model-free advantages with guaranteed fixed-time convergence properties.

Acknowledgements This research is supported by National Natural Science Foundation of China (No. U21A20485), International (Regional) Cooperation and Exchange (ICE) Projects of the National Natural Science Foundation of China (NSFC) (No. 62311540022), and China Postdoctoral Science Foundation (Project ID: 2024M762602).

Author contributions Junkai Tan: Conceptualization, Methodology, Writing, Visualization. Shuangsi Xue: Conceptualization, Methodology, Resources, Funding acquisition. Tianse Niu: Conceptualization, Resources, Visualization. Kai Qu: Validation, Investigation, Visualization. Hui Cao: Supervision, Funding acquisition, Project administration. Badong Chen: Supervision, Funding acquisition, Project administration.

Data Availability Statement No datasets were generated or analysed during the current study.

Declarations

Conflict of interest The authors declare that they have no Conflict of interest.

Conflict of interest The authors declare no Conflict of interest.

References

1. An, T., Zhu, X., Ma, B., Jiang, H., Dong, B.: Hierarchical approximate optimal interaction control of human-centered modular robot manipulator systems: a Stackelberg differential game-based approach. Neurocomputing **585**, 127573 (2024). https://doi.org/10.1016/j.neucom.2024.127573
2. Huang, D., Huang, T., Qin, N., Li, Y., Yang, Y.: Finite-time control for a UAV system based on finite-time disturbance observer. Aerosp. Sci. Technol. **129**, 107825 (2022). https://doi.org/10.1016/j.ast.2022.107825
3. Zhang, K., Zhang, Z.X., Xie, X.P., Rubio, J.D.J.: An unknown multiplayer nonzero-sum game: prescribed-time dynamic event-triggered control via adaptive dynamic programming. IEEE Trans. Autom. Sci. Eng. (2024). https://doi.org/10.1109/TASE.2024.3484412
4. Vahidi-Moghaddam, A., Mazouchi, M., Modares, H.: Memory-augmented system identification with finite-time convergence. IEEE Control Syst. Lett. **5**(2), 571–576 (2021). https://doi.org/10.1109/LCSYS.2020.3004423
5. Dong, B., Zhu, X., An, T., Jiang, H., Ma, B.: Barrier-critic-disturbance approximate optimal control of nonzero-sum differential games for modular robot manipulators. Neural Netw. **181**, 106880 (2025). https://doi.org/10.1016/j.neunet.2024.106880
6. He, Z., Shen, J., Zhang, Z.: Practical fixed-time tracking control of quadrotor unmanned aerial vehicles with input saturation. Asian Journal of Control https://doi.org/10.1002/asjc.3350
7. Tatari, F., Panayiotou, C., Polycarpou, M.: Nonlinear Discrete-time System Identification without Persistence of Excitation: Finite-time Concurrent Learning Methods (2022). https://doi.org/10.48550/arXiv.2112.07765
8. Huang, J., Wang, S., Wu, Z.: Robust Stackelberg Differential Game With Model Uncertainty. IEEE Trans. Autom. Control **67**(7), 3363–3380 (2022). https://doi.org/10.1109/TAC.2021.3097549
9. Li, M., Qin, J., Li, J., Liu, Q., Shi, Y., Kang, Y.: Game-based approximate optimal motion planning for safe human-swarm interaction. IEEE Trans. Cybernet. (2023). https://doi.org/10.1109/TCYB.2023.3340659
10. Zhang, Y., Zhang, P., Wang, X., Song, F., Li, C., Hao, J.: An open loop Stackelberg solution to optimal strategy for UAV pursuit-evasion game. Aerosp. Sci. Technol. **129**, 107840 (2022). https://doi.org/10.1016/j.ast.2022.107840
11. Tan, J., Wang, J., Xue, S., Cao, H., Li, H., Guo, Z.: Human-machine shared stabilization control based on safe adaptive dynamic programming with bounded rationality. Int. J. Robust Nonlinear Control (2025). https://doi.org/10.1002/rnc.7931
12. Kamalapurkar, R., Andrews, L., Walters, P., Dixon, W.E.: Model-based reinforcement learning for infinite-horizon approximate optimal tracking. IEEE Trans. Neural Netw. Learn. Syst. **28**(3), 753–758 (2017). https://doi.org/10.1109/TNNLS.2015.2511658
13. Abu-Khalaf, M., Lewis, F.L., Huang, J.: Neurodynamic programming and zero-sum games for constrained control systems. IEEE Trans. Neural Netw. **19**(7), 1243–1252 (2008). https://doi.org/10.1109/TNN.2008.2000204
14. Al-Tamimi, A., Lewis, F.L., Abu-Khalaf, M.: Model-free Q-learning designs for linear discrete-time zero-sum games with application to H-infinity control. Automatica **43**(3), 473–481 (2007). https://doi.org/10.1016/j.automatica.2006.09.019
15. Li, M., Qin, J., Freris, N.M., Ho, D.W.C.: Multiplayer Stackelberg-nash game for nonlinear system via value iteration-based integral reinforcement learning. IEEE Trans. Neural Netw. Learn. Syst. **33**(4), 1429–1440 (2022). https://doi.org/10.1109/TNNLS.2020.3042331
16. Li, M., Qin, J., Ma, Q., Zheng, W.X., Kang, Y.: Hierarchical optimal synchronization for linear systems via reinforcement learning: a Stackelberg-nash game perspective. IEEE Trans. Neural Netw. Learn. Syst. **32**(4), 1600–1611 (2021). https://doi.org/10.1109/TNNLS.2020.2985738
17. Ming, Z., Zhang, H., Li, Y., Liang, Y.: Mixed H_2/H_∞ control for nonlinear closed-loop stackelberg games with application to power systems. IEEE Trans. Autom. Sci. Eng. **21**(1), 69–77 (2024). https://doi.org/10.1109/TASE.2022.3216733
18. Li, Y., Yang, T., Tong, S.: Adaptive neural networks finite-time optimal control for a class of nonlinear systems. IEEE Trans. Neural Netw. Learn. Syst. **31**(11), 4451–4460 (2020). https://doi.org/10.1109/TNNLS.2019.2955438
19. Yue, H., Xia, J., Zhang, J., Park, J.H., Xie, X.: Event-based adaptive fixed-time optimal control for saturated fault-tolerant nonlinear multiagent systems via reinforcement learning algorithm. Neural Netw. **183**, 106952 (2025). https://doi.org/10.1016/j.neunet.2024.106952
20. Pita, J., Jain, M., Tambe, M., Ordóñez, F., Kraus, S.: Robust solutions to Stackelberg games: addressing bounded rationality and limited observations in human cognition. Artif.

Intell. **174**(15), 1142–1171 (2010). https://doi.org/10.1016/j.artint.2010.07.002

21. Lin, M., Zhao, B., Liu, D.: Event-Triggered Robust Adaptive Dynamic Programming for Multiplayer Stackelberg-Nash Games of Uncertain Nonlinear Systems. IEEE Trans. Cybernet. **54**(1), 273–286 (2024). https://doi.org/10.1109/TCYB.2023.3251653
22. Tan, L.N., Tran, H.T., Tran, T.T.: Event-triggered observers and distributed H∞ control of physically interconnected nonholonomic mechanical agents in harsh conditions. IEEE Trans. Syst. Man Cybernet. Syst. **52**(12), 7871–7884 (2022). https://doi.org/10.1109/TSMC.2022.3177043. (https://ieeexplore.ieee.org/document/9786038/)
23. Tan, L.N., Pham, T.C.: Optimal tracking control for PMSM with partially unknown dynamics, saturation voltages, torque, and voltage disturbances. IEEE Trans. Ind. Electron. **69**(4), 3481–3491 (2022). https://doi.org/10.1109/TIE.2021.3075892
24. Tan, J., Xue, S., Guo, Z., Li, H., Cao, H., Chen, B.: Data-driven optimal shared control of unmanned aerial vehicles. Neurocomputing **622**, 129428 (2025). https://doi.org/10.1016/j.neucom.2025.129428
25. Zhang, L., Chen, Y.: Finite-time adaptive dynamic programming for affine-form nonlinear systems. IEEE Trans. Neural Netw. Learn. Syst. (2023). https://doi.org/10.1109/TNNLS.2023.3337387
26. Wang, P., Yu, C., Lv, M., Cao, J.: Adaptive fixed-time optimal formation control for uncertain nonlinear multiagent systems using reinforcement learning. IEEE Trans. Netw. Sci. Eng. **11**(2), 1729–1743 (2024). https://doi.org/10.1109/TNSE.2023.3330266
27. Li, S., Ding, L., Zheng, M., Liu, Z., Li, X., Yang, H., Gao, H., Deng, Z.: NN-based reinforcement learning optimal control for inequality-constrained nonlinear discrete-time systems with disturbances. IEEE Trans. Neural Netw. Learn. Syst. **35**(11), 15507–15516 (2024). https://doi.org/10.1109/TNNLS.2023.3287881
28. Tan, J., Xue, S., Li, H., Guo, Z., Cao, H., Li, D.: Prescribed performance robust approximate optimal tracking control via Stackelberg game. IEEE Trans. Autom. Sci. Eng. (2025). https://doi.org/10.1109/TASE.2025.3549114
29. Dao, P.N., Nguyen, V.Q., Duc, H.A.N.: Nonlinear RISE based integral reinforcement learning algorithms for perturbed Bilateral Teleoperators with variable time delay. Neurocomputing **605**, 128355 (2024). https://doi.org/10.1016/j.neucom.2024.128355
30. Tan, J., Xue, S., Li, H., Cao, H., Li, D.: Safe Stabilization Control for Interconnected Virtual-Real Systems via Model-based Reinforcement Learning. In: 2024 14th Asian Control Conference (ASCC), pp. 605–610 (2024)
31. Modares, H., Lewis, F.L.: Optimal tracking control of nonlinear partially-unknown constrained-input systems using integral reinforcement learning. Automatica **50**(7), 1780–1792 (2014). https://doi.org/10.1016/j.automatica.2014.05.011
32. Modares, H., Lewis, F.L., Naghibi-Sistani, M.B.: Integral reinforcement learning and experience replay for adaptive optimal control of partially-unknown constrained-input continuous-time systems. Automatica **50**(1), 193–202 (2014). https://doi.org/10.1016/j.automatica.2013.09.043
33. Dao, P.N., Phung, M.H.: Nonlinear robust integral based actor-critic reinforcement learning control for a perturbed three-wheeled mobile robot with mecanum wheels. Comput. Electr. Eng. **121**, 109870 (2025). https://doi.org/10.1016/j.compeleceng.2024.109870
34. Wei, Z., Du, J.: Reinforcement learning-based optimal trajectory tracking control of surface vessels under input saturations. Int. J. Robust Nonlinear Control **33**(6), 3807–3825 (2023). https://doi.org/10.1002/rnc.6597
35. Tatari, F., Modares, H., Panayiotou, C., Polycarpou, M.: Finite-time distributed identification for nonlinear interconnected systems. IEEE/CAA J. Autom. Sin. **9**(7), 1188–1199 (2022). https://doi.org/10.1109/JAS.2022.105683
36. Tatari, F., Mazouchi, M., Modares, H.: Fixed-time system identification using concurrent learning. IEEE Trans. Neural Netw. Learn. Syst. **34**(8), 4892–4902 (2023). https://doi.org/10.1109/TNNLS.2021.3125145
37. Li, D., Ge, S., Lee, T.: Fixed-time-synchronized consensus control of multi-agent systems. IEEE Trans. Control Netw. Syst. (2020). https://doi.org/10.1109/TCNS.2020.3034523
38. Li, D., Ge, S., Lee, T.: Simultaneous arrival to origin convergence: sliding-mode control through the norm-normalized sign function. IEEE Trans. Autom. Control (2021). https://doi.org/10.1109/TAC.2021.3069816
39. Tan, J., Xue, S., Cao, H., Ge, S.S.: Human-AI interactive optimized shared control. J. Autom. Intell. (2025). https://doi.org/10.1016/j.jai.2025.01.001
40. Kamalapurkar, R., Walters, P., Dixon, W.E.: Model-based reinforcement learning for approximate optimal regulation. Automatica **64**, 94–104 (2016). https://doi.org/10.1016/j.automatica.2015.10.039
41. Mu, C., Wang, K., Zhang, Q., Zhao, D.: Hierarchical optimal control for input-affine nonlinear systems through the formulation of Stackelberg game. Inf. Sci. **517**, 1–17 (2020). https://doi.org/10.1016/j.ins.2019.12.078
42. Li, D., Ge, S., He, W., Ma, G., Xie, L.: Multilayer formation control of multi-agent systems. Automatica **109**, 108558 (2019). https://doi.org/10.1016/j.automatica.2019.108558
43. Liu, Y., Li, H., Lu, R., Zuo, Z., Li, X.: An overview of finite/fixed-time control and its application in engineering systems. IEEE/CAA J. Autom. Sin. **9**(12), 2106–2120 (2022). https://doi.org/10.1109/JAS.2022.105413
44. Tatari, F., Modares, H.: Deterministic and stochastic fixed-time stability of discrete-time autonomous systems. IEEE/CAA J. Autom. Sin. **10**(4), 945–956 (2023). https://doi.org/10.1109/JAS.2023.123405
45. Tatari, F., Niknejad, N., Modares, H.: Discrete-time nonlinear system identification: a fixed-time concurrent learning approach. IEEE Trans. Syst. Man Cybernet. Syst. (2024). https://doi.org/10.1109/TSMC.2024.3508267
46. Zhang, Z., Zhang, K., Xie, X., Stojanovic, V.: ADP-based prescribed-time control for nonlinear time-varying delay systems with uncertain parameters. IEEE Trans. Autom. Sci. Eng. (2024). https://doi.org/10.1109/TASE.2024.3389020
47. Tan, J., Xue, S., Guan, Q., Qu, K., Cao, H.: Finite-time safe reinforcement learning control of multi-player nonzero-sum game for quadcopter systems. Inf. Sci. (2025). https://doi.org/10.1016/j.ins.2025.122117
48. Zhang, Z.X., Zhang, K., Xie, X.P., Sun, J.Y.: Fixed-time zero-sum pursuit-evasion game control of multi-satellite via adaptive dynamic programming. IEEE Trans. Aerosp.

Journal: **11071** MS: **11235** TYPESET DISK LE CP Disp.:**2025/5/2** Pages: **21** Layout: **Medium**

Electron. Syst. (2024). https://doi.org/10.1109/TAES.2024.3351810
49. Tan, J., Xue, S., Guan, Q., Niu, T., Cao, H., Chen, B.: Unmanned aerial-ground vehicle finite-time docking control via pursuit-evasion games. Nonlinear Dyn. (2025). https://doi.org/10.1007/s11071-025-11021-6
50. Tan, L.N., Gia, D.L.: ADP-Based H_∞ optimal decoupled control of single-wheel robots with physically coupling effects, input constraints, and disturbances. IEEE Trans. Ind. Electron. **71**(7), 7445–7454 (2024). https://doi.org/10.1109/TIE.2023.3301537
51. Kamalapurkar, R., Dinh, H., Bhasin, S., Dixon, W.E.: Approximate optimal trajectory tracking for continuous-time nonlinear systems. Automatica **51**, 40–48 (2015). https://doi.org/10.1016/j.automatica.2014.10.103
52. Perrusquía, A.: A complementary learning approach for expertise transference of human-optimized controllers. Neural Netw. **145**, 33–41 (2022). https://doi.org/10.1016/j.neunet.2021.10.009
53. Van Der Schaft, A.: L_2-gain analysis of nonlinear systems and nonlinear state-feedback H_∞ control. IEEE Trans. Autom. Control **37**(6), 770–784 (1992). https://doi.org/10.1109/9.256331
54. Nguyen Tan, L.: Distributed optimal control for nonholonomic systems with input constraints and uncertain interconnections. Nonlinear Dyn. **93**(2), 801–817 (2018). https://doi.org/10.1007/s11071-018-4228-8
55. Le-Dung, N., Huynh-Lam, P., Hoang-Giap, N., Tan-Luy, N.: Event-triggered distributed robust optimal control of nonholonomic mobile agents with obstacle avoidance formation, input constraints and external disturbances. J. Franklin Inst. **360**(8), 5564–5587 (2023). https://doi.org/10.1016/j.jfranklin.2023.02.033
56. Polyakov, A.: Nonlinear feedback design for fixed-time stabilization of linear control systems. IEEE Trans. Autom. Control **57**(8), 2106–2110 (2012). https://doi.org/10.1109/TAC.2011.2179869
57. Filippov, A.F.: Differential Equations with Discontinuous Righthand Sides, *Mathematics and Its Applications*, vol. 18. Springer Netherlands, Dordrecht (1988). https://doi.org/10.1007/978-94-015-7793-9
58. Wang, D., Qiao, J.: Approximate neural optimal control with reinforcement learning for a torsional pendulum device. Neural Netw. **117**, 1–7 (2019). https://doi.org/10.1016/j.neunet.2019.04.026
59. Yu, S., Zhang, H., Ming, Z., Sun, J.: Adaptive optimal control via continuous-time Q-learning for stackelberg-nash games of uncertain nonlinear systems. IEEE Trans. Syst. Man Cybernet. Syst. **54**(7), 4461–4470 (2024). https://doi.org/10.1109/TSMC.2024.3382356
60. Zhang, L., Chen, Y.: Distributed finite-time ADP-based optimal secure control for complex interconnected systems under topology attacks. IEEE Trans. Syst. Man Cybernet. Syst. **54**(5), 2872–2883 (2024). https://doi.org/10.1109/TSMC.2024.3351909
61. Bhasin, S., Kamalapurkar, R., Johnson, M., Vamvoudakis, K.G., Lewis, F.L., Dixon, W.E.: A novel actor-critic-identifier architecture for approximate optimal control of uncertain nonlinear systems. Automatica **49**(1), 82–92 (2013). https://doi.org/10.1016/j.automatica.2012.09.019

Journal: **11071** MS: **11235** TYPESET DISK LE CP Disp.:**2025/5/2** Pages: **21** Layout: **Medium**